THE UNIVERSITY *of* EDINBURGH

Investigating the Structure and Dynamics of DNA with Fluorescence and Computational Techniques

Darren Andrew Smith

Submitted in total fulfilment of the requirements of the degree of Doctor of Philosophy under a jointly-awarded PhD programme

The University of Edinburgh

•

The University of Melbourne

2014

Produced on archival quality paper

Abstract

Nucleic acids, such as DNA, play an essential role in all known forms of life; however, despite their fundamental importance, there is still a significant lack of understanding surrounding their functional behaviour. This thesis explores the structure and dynamics of DNA by employing methods based on fluorescence and through the use of computational calculations.

Time-resolved fluorescence experiments have been performed on dinucleotides containing 2-aminopurine (2AP) in various alcohol-water mixtures. 2AP, a fluorescent analogue of the nucleobase adenine, has been used extensively to investigate nucleic acids because of its ability to be incorporated into their structures with minimal perturbation and its high sensitivity to its local environment. Direct solvent effects on 2AP were established through measurements on the free fluorophore. Analysis of the complex fluorescence decays associated with the dinucleotides was challenging but has provided insight into their conformational dynamics. Solvent polarity was found to play a significant role in determining both photophysical and conformational properties in these systems.

The complicated fluorescence decay of 2AP in nucleic acids highlights the need for accurate and unbiased analysis methods. Various time-resolved fluorescence analysis methods, including iterative reconvolution and the exponential series method, have been investigated with real and simulated data to obtain an overview of their benefits and limitations. The main outcome of the evaluation is that no single method is preferred in all situations and there is likely to be value in using a combination when there is ambiguity in the interpretation of the results.

Regardless of the analysis technique used, the parameterised description of the observed fluorescence decay is meaningless if the underlying physical model is unrealistic. The advance of computational methods has provided a new means to rigorously test the viability of proposed models. Calculations have been performed at the M06-2X/6-31+G(d) level of theory to investigate the stability of 2AP-containing dinucleotides in conformations similar to those observed in the double-helical structure of DNA. The results help to explain the similarity of the time-resolved fluorescence behaviour of 2AP in dinucleotide and DNA systems but also bring to light subtle differences that could perhaps account for experimental discrepancies.

The recent emergence of advanced optical microscopy techniques has offered the prospect of being able to directly visualise nucleic acid structure at the nanoscale but, unfortunately, limitations of existing labelling methods have hindered delivery of this potential. To address this issue, a novel strategy has been used to introduce reversible fluorescence photoswitching into DNA at high label density. Photophysical studies have implicated aggregation and energy-transfer as possible quenching mechanisms in this system, which could be detrimental to its future application. The reliability of fluorescence photoswitching was investigated at ensemble and single-molecule level and by performing optical lock-in detection imaging. These developments lay the foundations for improved and sequence-specific super-resolution microscopy of DNA, which could offer new insights into the 3D nanoscale structure of this remarkable biopolymer.

In summary, the work presented in this thesis outlines important observations and developments that have been made in the study of the structure and dynamics of nucleic acids.

Lay Summary

The overall aim of this thesis has been to explore the structure and dynamics of deoxyribonucleic acid; better known simply as DNA. Despite being fundamental to all known forms of life there is still a lack of knowledge concerning how DNA functions. The most common picture of DNA is the iconic double-helical structure first described by Watson and Crick back in the 1950s. Building blocks called nucleotides link together to create chain-like strands. Each nucleotide is constructed of three smaller parts: a sugar group, a phosphate group, and a base – of which there are four different types. The sugar and phosphate groups form a repeating structure that acts as the backbone to the DNA strand while the base (or rather sequence of bases) encodes genetic information: the instructions of life. The helical shape of double-stranded DNA is caused by base-stacking and base-pairing interactions. Although detailed, this idealised, static picture of DNA cannot explain how it actually performs its function within the cell. Given the importance of such processes to how life functions, there is a need to move towards a more dynamic understanding of DNA.

DNA is extremely complex and so the work presented in this thesis aims to offer progressive insight into its behaviour by studying model systems. There has been a particular focus on dinucleotides: small fragments of just two nucleotide units that provide a greatly simplified model of DNA by removing all but one neighbouring base. Experimental work has mainly been performed using fluorescence-based techniques. A molecule that absorbs light is promoted into a higher energy (excited) state and one of the ways it can lose this extra energy is to fluoresce; that is, emit light back out. The brightness (intensity) and colour (frequency) of the emitted light, as well as the delay between absorption and emission (measurable with time-resolved fluorescence techniques), can be used to gain insight into the fluorescent molecule and its environment. Labelling DNA with fluorescent probes is a very powerful interrogation method because probes can be sensitive to their local surroundings, reporting on conformational or environmental changes that occur, or can light-up structure, providing insight into the global architecture of DNA. The rapid advance of computational methods has offered a new way to investigate almost any conceivable system of interest. This has enabled complementary studies to be performed that help to elucidate experimental results.

In summary, fluorescence-based analytical methods, including the time-resolved spectroscopy of fluorescent base-analogue 2AP, have been used in combination with computational calculations to investigate the conformational properties of DNA that determine its functional behaviour. Complementary studies of data analysis methods have been performed to facilitate interpretation of the data, clarify results, and enhance the overall understanding of this complex system. A novel DNA substrate for advance fluorescence microscopy has also been developed. This type of material could offer the prospect of directly imaging the 3D nanoscale structure of DNA in cellular environments. The results offer new insights into the structure and dynamics of DNA.

Declaration

This thesis has been submitted for fulfilment of requirements of a Universitas 21 Joint PhD, which was hosted by the University of Edinburgh, United Kingdom (lead institution) and the University of Melbourne, Australia (partner institution).

This is to certify that that the work contained within has been composed by me and is entirely my own work, except where otherwise stated. I confirm that appropriate credit has been given within the thesis where reference has been made to the work of others.

Chapter 6 of this thesis has been based on work from the published paper entitled 'Reversible Fluorescence Photoswitching in DNA' (D. A. Smith, P. Holliger, and C. Flors, *J. Phys. Chem. B*, 2012, **116**, 10290–3, DOI: 10.1021/jp3056834). Notice of the use of this work will be presented appropriately before the relevant chapter.

No part of this thesis has been submitted for any other degree or professional qualification.

Signed:

_____

Date:

_____

Acknowledgements

This thesis would never have been completed without the help and support of many people.

Firstly, I express sincerest gratitude to my supervisors, Prof. Anita C. Jones and A/Prof. Trevor A. Smith, for giving me the opportunity to take on this challenging but thoroughly enjoyable project. I greatly appreciate the time they have spent explaining things to me, the guidance they have provided me, and the trust they afforded me.

Secondly, I am deeply obliged to Dr. Cristina Flors for showing me the nanoscopic realm, helping me produce my first publication, and teaching me how to be a scientist.

Thirdly, I thank Dr. Tanja van Mourik and Leo Holroyd of the University of St Andrews, United Kingdom, and Douglas M. Matje, Norbert O. Reich, and others at the University of California, United States of America, for productive collaboration; I can only hope my contribution has been as worthwhile to them as theirs has been to me.

Fourthly, I am greatly indebted to Dr. Emma Hooley and the rest of her family for making me feel so welcome so far from home; their kindness will never be forgotten.

Fifthly, I would like to show my appreciation to the following: the Jones Group, past and present, for teaching me many new things and for their friendship; Dr. Ben Robotham, Ben Morrison, and the other members of the Smith-Ghiggino consortium in Melbourne, for insightful discussion and most excellent company, particularly while indulging upon falafel wraps and chicken curry don; Dr. Xiao-Tao Hao, for his much needed assistance in the laser laboratory; Dr. Jöchen Arlt, for tirelessly replacing mirrors so I could perform my experiments; Dr. John White, for his guidance during PCR and ethanol precipitation procedures; Dr. David Rodgers, for leading me through the mysterious world of Gaussian; Edinburgh University Roller Hockey Club, for some great experiences and for providing an excellent distraction during frustrating times; and last but not least, the people unseen, those who provided my funding and those who keep the clock ticking smoothly.

Finally, I thank all of my friends and family for being there when I need them. I am especially grateful to my sister, Angela, for not letting me take myself too seriously and putting a smile on my (pie) face; and my brother, Gary, for his sagely advice and continual encouragement. Greatest thanks of all go to my mum, Morag, and dad, Douglas, for supporting me throughout my life and giving me the opportunity to do something I enjoy; I could not wish for more.

Table of Contents

Abstract **iii**

Lay Summary **v**

Declaration **vii**

Acknowledgements **ix**

Chapter 1: Introduction **1**

1.1 Purpose and Context of Thesis 3

1.2 Structure of Thesis 5

1.3 References 7

Chapter 2: Background and Theory **9**

2.1 Fluorescence Spectroscopy 11

2.1.1 Absorption (Excitation) *11*

2.1.2 Transition Dipole Moment *11*

2.1.3 Relaxation Processes *12*

2.1.4 Excited-State Dynamics *18*

2.2 Computational Chemistry 20

2.3 DNA Structure 25

2.3.1 Fundamentals of DNA Structure *25*

2.3.2 A Brief History of DNA Structure *34*

2.3.3 Fluorescent Nucleic Acid Base Analogues *39*

2.4 References 48

Chapter 3: Time-Resolved Fluorescence of 2AP-Containing Dinucleotides in Alcohol-Water Mixtures **57**

3.1 Introduction 59

3.2 Materials and Methods 66

3.2.1 Sample Preparation *66*

3.2.2 Steady-State Measurements *67*

3.2.3 Time-Correlated Single Photon Counting (TCSPC) *67*

3.3 Results and Discussion 73
3.3.1 Steady-State Spectroscopy of 2AP 73
3.3.2 Steady-State Spectroscopy of 2AP-N Dinucleotides 78
3.3.3 Time-Resolved Spectroscopy of 2AP 81
3.3.4 Time-Resolved Spectroscopy of 2AP-N Dinucleotides 86
3.3.5 Implications of the Solvent-Dependence of Dinucleotide Decay Parameters 100
3.4 Conclusions 102
3.4.1 2AP Fluorescence in Solution 102
3.4.2 2AP-Containing Dinucleotides in Water 102
3.4.3 2AP-Containing Dinucleotides in Alcohol-Water Mixtures 103
3.4.4 Outlook 104
3.5 References 106

Chapter 4: A Comparison of Analysis Techniques for Time-Resolved Fluorescence 109

4.1 Introduction 111
4.1.1 Deconvolution and an Ill-Posed Problem 111
4.1.2 Analysis Techniques for Time-Resolved Fluorescence 113
4.2 Methods 122
4.2.1 Experimental and Simulated Decays 122
4.2.2 Decay Analysis 123
4.3 Results and Discussion 126
4.3.1 Analysis of Simulated Decays 128
4.3.2 Analysis of Real Decays 137
4.3.3 Evaluation of Time-Resolved Fluorescence Analysis Methods 144
4.4 Conclusions 147
4.4.1 Improvements to Analysis Methods 147
4.4.2 Outlook 149
4.5 References 150

Chapter 5: A Computational Study of 2AP-Containing Dinucleotides 153

5.1 Introduction 155
5.1.1 A Computational Approach to DNA Structure 156
5.1.2 Structures Investigated in this Study 161

5.2 Methods 163
5.2.1 Molecular Structure 163
5.2.2 Computational Calculations 165
5.2.3 Analysis of Stacking Geometry 166
5.3 Results and Discussion 168
5.3.1 Anionic Dinucleotide or Capped Phosphate Group? 168
5.3.2 Base-Stacked Structures of 2AP-N Dinucleotides in Water 172
5.3.3 Comparison of 2AP and Adenine Dinucleotides 177
5.3.4 Structures of 2AP|N Dimers in Water 180
5.3.5 Structure and Interactions of the Sugar-Phosphate Backbone 185
5.3.6 Effects of PCM Solvent: Structures in Water and the Gas-Phase 189
5.3.7 Open Structures and Base-Flipping 197
5.3.8 Noncanonical Structures 199
5.4 Conclusions 204
5.4.1 Dinucleotide Base-Stacking is Similar to DNA Duplex Conformation 204
5.4.2 Structure and 2AP Fluorescence Lifetime 205
5.4.3 2AP as an Adenine Analogue 206
5.4.4 The Sugar-Phosphate Backbone Influences Conformation 206
5.4.5 PCM Solvent Effects 208
5.4.6 Caveats and Possible Future Experiments 209
5.4.7 Outlook 212
5.5 References 214
Chapter 6: Reversible Fluorescence Photoswitching in DNA 219
6.1 Introduction 221
6.1.1 Advantages of Microscopy 222
6.1.2 Microscopy Techniques 223
6.1.3 Spatial Resolution and the Diffraction Limit 224
6.1.4 Optical Control: Photoswitching 229
6.1.5 Photoswitching and Super-Resolution 231
6.1.6 Super-Resolution Imaging and DNA 233
6.1.7 Contrast and Optical Lock-In Detection Microscopy 234
6.1.8 CyDNA 236

6.1.9 CyDNA and OLID *239*
6.2 Materials and Methods 241
6.2.1 Synthesis of CyDNA *241*
6.2.2 Photoswitchable CyDNA (psCyDNA) *242*
6.2.3 Preparation of CyDNA-Coated Beads *243*
6.2.4 Photophysical Studies *243*
6.2.5 Microscopy *243*
6.3 Results and Discussion 246
6.3.1 Confirmation of PCR Product by Agarose Gel Electrophoresis *246*
6.3.2 Photophysics of CyDNA *247*
6.3.3 Photoswitching of CyDNA: Ensemble and Single-Molecule Measurements *251*
6.3.4 Optical Lock-In Detection Microscopy *254*
6.3.5 Super-Resolution Imaging of Cy3/5DNA *260*
6.4 Conclusions 261
6.5 References 264

Chapter 7: Conclusions **271**

Chapter 1: Introduction

To condense fact from the vapor of nuance.

– Neal Stephenson (Juanita), *Snow Crash*

1.1 Purpose and Context of Thesis

Deoxyribonucleic acid (DNA) is of fundamental importance to our existence yet there still remains a significant shortfall in knowledge concerning its functional behaviour. In particular, it has become evident that the static and idealised model of the double helix – first devised by Watson and Crick in the 1950s – is insufficient to explain the observed phenomena in which DNA plays a central role. This thesis outlines important observations and developments that have been made in the study of the structure and dynamics of DNA. New insights have been established through the use of methods based on fluorescence and by performing computational calculations.

Fluorescence techniques are ideally suited to probing conformational dynamics due to their inherent sensitivity and selectivity; however, as the natural nucleic bases are essentially non-fluorescent, a suitable probe has to be introduced to be able to study DNA by these means. Here, steady-state and time-resolved fluorescence spectroscopy has been enabled by the use of 2-aminopurine (2AP), which has highly desirable photophysical and physiological properties. 2AP is a fluorescent analogue of the nucleobase adenine and can be incorporated into nucleic acid constructs with minimal perturbation to the native structure. It addition, its time-resolved fluorescence decay is highly sensitivity to changes in its local environment, particularly when base-stacking interactions are present.

Computational approaches offer a unique perspective of complex systems because of the ability to control the precise details of the simulated system. Models that are used for time-resolved fluorescence decays are often influenced or justified by predictions of dynamic conformational behaviour in the system; however, as is the case with 2AP in nucleic acids, these assumptions are often based on preconceived ideas about molecular structure, which are typically inferred from static, and potentially perturbed, crystallographic data. The advance of computational methods has provided a means to rigorously test the viability of proposed physical models.

Microscopy has historically provided the most intuitive understanding of biological systems due to its ability to directly visualise the microscopic realm of the cell but, unfortunately, the scale of DNA (as well as many other biological components) goes far beyond the resolution threshold for conventional imaging methodologies. In addition, the fragile nature of biological systems means that the introduction of exogenous components can cause significant perturbation to the native state; there is therefore a need for methods that can produce high quality images at very low probe

concentrations. Over the past decade or so the rapid advance of new fluorescence-based imaging strategies, including those collectively known as super-resolution techniques, has offered the prospect of acquiring unprecedented detail of the cellular world; however, there are still significant barriers that need to be addressed before the full potential of these techniques is delivered, especially for systems where DNA is the target species.

Current strategies to label DNA for fluorescence microscopy are limited by a number of factors. Fluorescent base analogues, such as 2AP, have proven very useful for investigating changes in the local structure and dynamics of nucleic acid systems but a single fluorophore is incapable of providing information about the global properties of large macromolecules. Incorporation of a large number of labels at high density is required to obtain a broader picture of the system but this is a significant challenge and there are few methods currently available to achieve this goal.

Synthesising DNA substrates with base analogues is difficult and potentially expensive. Additionally, while a single mutation may be easily accommodated by the duplex, there is a risk that replacing many of the natural bases with modified ones will cause significant perturbation to the overall structure, which may disrupt normal function of the system. Dyes that can intercalate or bind in the grooves of DNA offer the potential to allow dense labelling but their distribution is random, meaning coverage might be heterogeneous, and they can also cause structural changes to occur. Another option is to use DNA-associated proteins for imaging. These proteins can be non-specific, where they interact stochastically with the DNA, or can target particular recognition sequences, which can help to isolate a region-of-interest in the system. Proteins can be easily labelled or even reengineered to fluoresce but it is also possible to use intrinsic biological fluorophores, such as tryptophan, for detection. This strategy is appealing but it is limited by the fact that protein binding is at equilibrium and so large concentrations are required to ensure adequate coverage of the DNA; this can lead to large background signals. Additionally, proteins can by very bulky and may be unable to reach a target located in a densely packed environment, such as that found within the cell.

As well as obtaining high label content it is also necessary to use a suitable label for imaging. While providing extensive spectroscopic information, the ultraviolet (UV) light absorbed and emitted by 2AP and intrinsic biological fluorophores is of limited use in fluorescence microscopy. In addition to the fact excitation energies are high, potentially damaging the system, using short wavelength UV light is hindered by

significant scattering, absorption by common (glass) optics, and also inefficient detection. There is therefore a need to move to fluorescent labels that are characterised by spectra in the visible and infrared (IR) regions. Furthermore, the DNA system that is to be imaged must also be capable of reversible fluorescence photoswitching, which is an optical control method that underpins many of the advanced microscopy techniques that have been developed to overcome conventional resolution limits.

1.2 Structure of Thesis

Now that the underlying context has been addressed, a thorough outline of the structure of this thesis will be presented. It should be noted that appendices and other supplementary information, such as analysis code and demonstration movies, will not be included with the hardbound copy of the thesis but can be found with the electronic version. This thesis was composed in Microsoft Word (2010)[1] and, unless otherwise stated, all figures were created using a combination of Microsoft Excel (2010),[2] Microsoft PowerPoint (2010),[3] ChemDraw,[4] and MATLAB.[5]

Chapter 2 will briefly introduce the fundamental background knowledge and theoretical considerations that underlie fluorescence spectroscopy, computational chemistry, DNA structure, and the photophysical properties of 2AP in nucleic acid constructs. This chapter will not extensively cover all aspects of the work that has been performed; more specific details, including methodologies, will be presented later in the relevant chapter.

Chapter 3 reports the use of time-resolved fluorescence experiments to investigate 2AP-containing dinucleotides in various alcohol-water mixtures. The use of the dinucleotide system offered a simplified model of DNA and allowed base-specific effects to be isolated. Direct solvent effects on 2AP are established by making measurements on the free fluorophore. A comparative approach is taken to analyse the more complicated fluorescence decay associated with the dinucleotides.

Chapter 4 explores various time-resolved fluorescence analysis methods. An overview of their benefits and limitations is obtained by using real and simulated data. Methods studied include iterative reconvolution, the exponential series method, and a recently proposed method based on the ideas of compressed sensing. A deconvolution approach that is founded on the process of simulated annealing is also investigated.

Chapter 5 is concerned with the computational investigation of the stability of 2AP-containing dinucleotides that are in conformations similar to those observed in the canonical, double-helical structure of DNA. The structures are evaluated by performing

calculations at the M06-2X/6-31+G(d) level of theory. Analogous nucleobase dimer structures are also considered to assess the influence of the sugar-phosphate backbone, which is often omitted in studies to reduce computational cost. The work presented in Chapter 5 greatly benefited from collaboration with Dr. Tanja van Mourik and Leo Holroyd from the University of St Andrews. Of particular note is that the methodology that was adopted for calculations was heavily dependent on their expertise and experience. Where appropriate, their contribution to other aspects of the study will be highlighted.

Chapter 6 reports on a novel strategy to introduce reversible fluorescence photoswitching into DNA at high label density. The recent development of a modified DNA polymerase that has the capacity to efficiently incorporate Cy3- and Cy5-labelled cytidine analogues during polymerase chain reaction has enabled sequence-specific labelling of double-stranded DNA. The resulting substrate, known as CyDNA, can reach around 1.3 kbp in length with hundreds of fluorophores decorating each strand. Photophysical studies are performed to establish possible quenching mechanisms in this system. The reversible fluorescence photoswitching capability of the Cy3-Cy5 pair is introduced into CyDNA to produce a novel class of nanomaterial that holds promise for advanced microscopy applications. The reliability of fluorescence photoswitching in CyDNA is investigated at ensemble and single-molecule level and by performing optical lock-in detection imaging. This chapter is based on results that have been published.

Finally, Chapter 7 provides an overview of the conclusions drawn from the earlier chapters. Some potential future experiments are also highlighted.

1.3 References

1. Microsoft Word (2010), *Version 14.0*, Microsoft Corporation, Microsoft Redmond Campus, Redmond, Washington, United States of America, 2010.

2. Microsoft Excel (2010), *Version 14.0*, Microsoft Corporation, Microsoft Redmond Campus, Redmond, Washington, United States of America, 2010.

3. Microsoft PowerPoint (2010), *Version 14.0*, Microsoft Corporation, Microsoft Redmond Campus, Redmond, Washington, United States of America, 2010.

4. ChemDraw, *Version 14.0 Std*, CambridgeSoft Corporation, a subsidiary of PerkinElmer, Inc., Waltham, Massachusetts, United States of America, 2014.

5. MATLAB, *Version 8.2.0 (R2013b)*, The MathWorks, Inc., Natick, Massachusetts, United States of America, 2013.

Chapter 2: Background and Theory

In the beginning, there was nothing, which exploded.

– Terry Pratchett, *Lords and Ladies*

This chapter outlines some of the common principles that underlie the results presented throughout this thesis. Basic concepts related to fluorescence spectroscopy and computational chemistry will be introduced as well as some background knowledge concerning the general structural features of DNA. To conclude, there will be a short section about fluorescent nucleic acid base analogues. Particular attention will be paid to the properties of 2-aminopurine (2AP) that make it a highly desirable probe for systems involving nucleic acids.

2.1 Fluorescence Spectroscopy

This section will highlight some of fundamental aspects of fluorescence spectroscopy. The content is mainly based on the work of Hollas,[1] Atkins and de Paula,[2] Turro,[3] and Lakowicz.[4]

2.1.1 Absorption (Excitation)

A molecule in an initial state of energy E_i can absorb a photon to access some excited state of higher energy E_f. Figure 2.1 shows a schematic of this process.

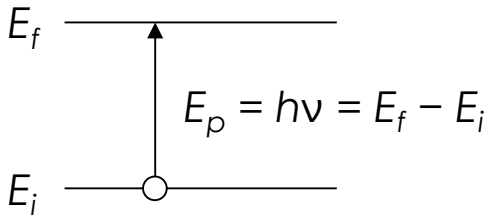

Figure 2.1 – Absorption of a photon can promote an electron into a higher energy excited state. The photon energy, E_p, is given by Planck's constant, h, and the frequency of the light, ν.

The electronic and vibrational (vibronic) level of the molecule that is accessed in this process is dependent on the energy of the photon, E_p, that is absorbed and the energy of the state which is initially occupied – most commonly the ground (electronic) state. Absorption is a fundamental interaction that occurs between light and matter and is possible due to the coupling of electron motion with the oscillating electric field of the incident light.

2.1.2 Transition Dipole Moment

The probability of a transition occurring between two electronic states is dependent on the transition dipole moment (TDM, $\mu_{i\to f}$). The TDM is related to the

overlap of the wavefunctions of the initial state, ψ_i, and the final state, ψ_f, and can be calculated using the electric dipole moment operator, $\hat{\mu}$;

$$\mu_{i\to f} = \int \psi_i^* \cdot \hat{\mu} \cdot \psi_f d\tau, \qquad 2.1$$

where dτ is the volume element and denotes that the integration is performed over all space. If the TDM is zero then a transition is *forbidden* but if it is non-zero then the transition is *allowed*.

The wavefunction holds the full, quantum mechanical description of the state of a system. In principle, any property of a system can be calculated using the wavefunction; however, apart from extremely simple systems, it is generally impossible to obtain an analytical representation that accurately describes its complete form. This issue can be addressed to some extent by splitting the overall wavefunction into various parts that are defined by independent variables. For instance, the Born-Oppenheimer approximation allows electronic and nuclear parts of the wavefunction to be uncoupled from each other by assuming that, due to vast difference in mass, electrons instantaneously accommodate any change in nuclear position (due to translational, vibrational, or rotational motion). From an alternative perspective, this can also be taken to mean that the nuclei are essentially static relative to electron motion. This is the foundation of the Franck-Condon principle, which approximates electronic transitions as vertical (orthogonal) transitions relative to nuclear coordinates.

2.1.3 Relaxation Processes

A molecule in an excited state is no longer in its most stable configuration and so there is a driving force towards returning to the ground state. This *relaxation* process can be achieved in a number of different ways.

2.1.3.1 Fluorescence and Phosphorescence

In the opposite manner to absorption, it is possible that the molecule may relax *via* emission of a photon. The general term for this process is luminescence; however, it can be further categorised as either fluorescence or phosphorescence depending on the electron spin properties of the transition. Since the spin-states of an electron are orthogonal to each other (generally referred to as being either *spin up* or *spin down*), the relative orientation of electron spin, which is characterised by spin multiplicity, must be maintained for radiative emission to be allowed. Spin multiplicity is calculated

as $2S + 1$, where $S = \sum m_s$; that is, the sum of all spin quantum numbers, m_s. Since electrons are fermions with $m_s = \pm ½$, the value of S can be conveniently calculated as half the number of unpaired electrons. When all electrons are paired ($S = 0$) the spin multiplicity has a value of 1 and the electronic state is a called a singlet state. As a consequence of the Pauli exclusion principle, which states that no two fermions can share an identical set of quantum numbers (including spin), the ground state is generally of singlet character to maximise occupancy of the lowest energy orbitals available. When there are two unpaired electrons ($S = 1$), which might occur if one electron is promoted into a higher energy orbital, then the spin multiplicity has a value of 3 and the electronic state is called a triplet state. Following from this, fluorescence, which is defined as a singlet-singlet transition, is an allowed process but phosphorescence, which is defined as a triplet-singlet transition, is a forbidden process. The rate of transition between two states is related to the square of the TDM. Thus, allowed transitions occur much more frequently than forbidden ones. Typically, the lifetime (the inverse of the transition rate) that characterises a fluorescence transition is on the picosecond (ps) to nanosecond (ns) timescale; this is significantly shorter than the millisecond (ms) to second (s) timescale usually observed for phosphorescence. It should be noted that these lifetimes characterise the occurrence of the process rather the electronic transition itself, which happens on the femtosecond (fs) timescale in both instances.

Spin-Orbit Coupling

If the system does not have a (symmetric) central potential then the spin quantum number becomes poorly defined; consequently, the requirement for maintaining spin multiplicity during an electronic transition is less rigorous. Spin-orbit coupling, which is a magnetic interaction between electrons and nuclei, can break the symmetry of a potential and, thus, can allow triplet-singlet (singlet-triplet) transitions to occur. Heavy-atoms induce a greater spin-orbit coupling (due to a larger nuclear charge) and so increase the likelihood of such a transition.

2.1.3.2 Vibrational Relaxation and Kasha's Rule

Nonradiative relaxation processes compete with fluorescence and phosphorescence. A molecule in an excited vibrational state may undergo vibrational relaxation. This is where collisions with neighbouring species vibrationally cool the excited molecule. If the collision rate is high (such as would be expected in a condensed

phase) then vibrational relaxation can be very efficient and the excited species will only populate vibrational states which are thermally accessible. The occupancy of each state is therefore determined by the Boltzmann distribution. It is often assumed that only the ground vibrational state will be occupied in electronically excited states. This assumption is known as Kasha's Rule and means that the observed emission spectrum is generally independent of the excitation wavelength.

2.1.3.3 Internal Conversion and Intersystem Crossing

Internal conversion (IC) and intersystem crossing (ISC) are also important nonradiative processes. IC and ISC are transitions between different vibronic states which have the same energy. IC is a transition between electronic states with same spin multiplicity while ISC involves transition to an electronic state with a different spin multiplicity. ISC, like phosphorescence, is therefore spin forbidden and requires spin-orbit coupling to occur.

IC and ISC rates depend on the overlap of the vibrational wavefunctions of the states associated with the transition of interest. This overlap is known as the Franck-Condon factor. Due to an increase in the density of vibronic states, the transition probability is greater for higher energy electronic states. A greater Franck-Condon factor is also achieved when the potential energy curves of the two electronic states involved overlap. This is because vibrational wavefunctions, with the exception of the ground vibrational state, have their maxima near the edge of electronic potential surface.

Figure 2.2 shows a simplified Jablonski diagram of the competitive relaxation processes discussed above. S_0, S_1, and S_2 are singlet ground, first excited, and second excited electronic states, respectively. T_1 represents the triplet first excited state where two electrons are unpaired and have the same spin orientation. Triplet states have lower energy than equivalent singlet states due to the fact they have greater (spin) angular momentum, which reduces the effect of electron repulsion by allowing the electrons to better avoid each other. Each electronic state has many associated vibrational energy levels (thin lines).

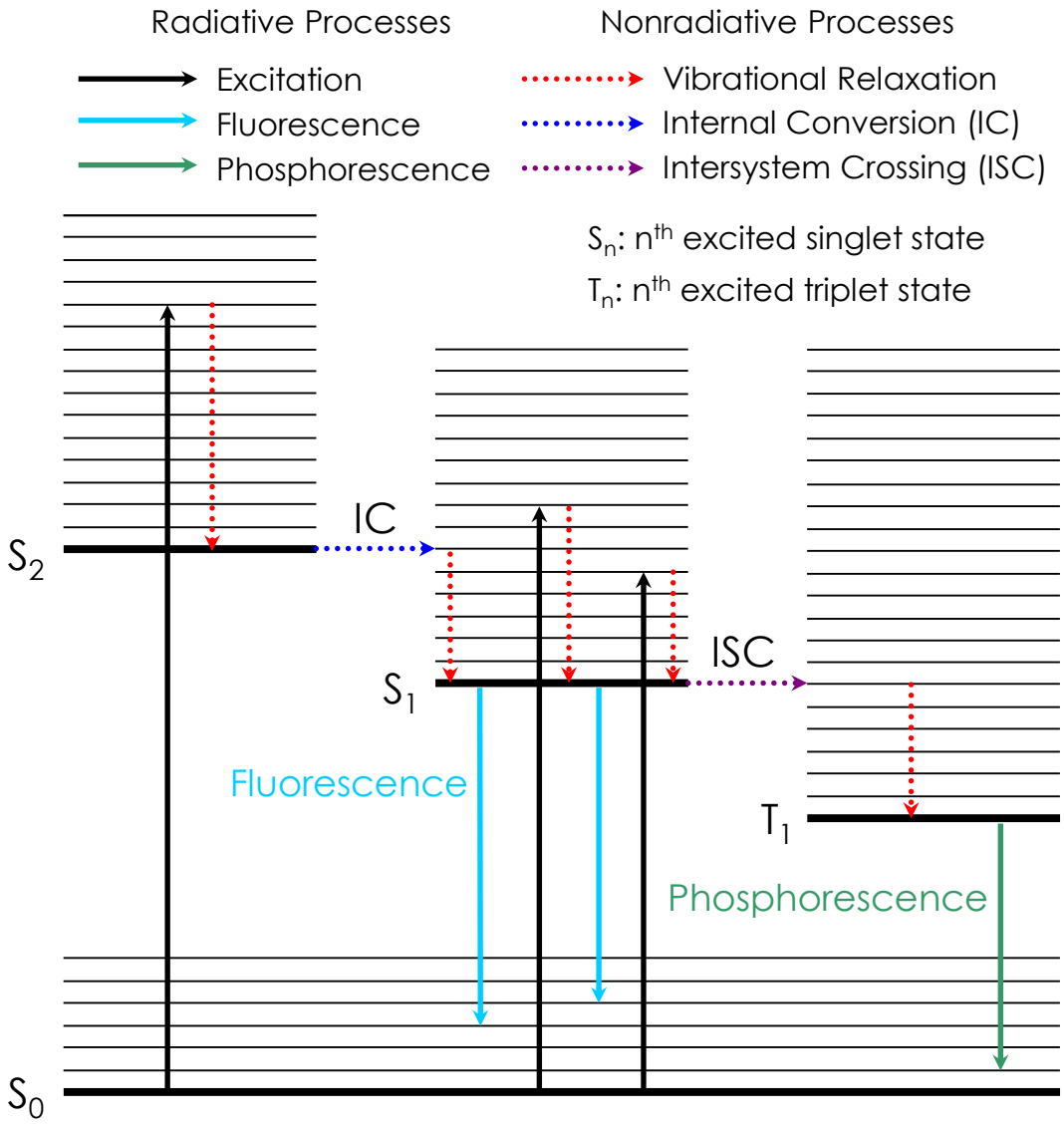

Figure 2.2 – A Jablonski diagram showing the vibronic transitions that may be possible in a molecule. Once in an excited state, a molecule may relax back down to its ground state, S_0, through many different pathways. Thin lines represent vibronic states. Heavy lines are vibrationally cold states. Upward arrows represent excitation (*via* absorption), downward arrows represent relaxation processes and horizontal lines represent isoenergetic processes.

In addition to the processes that have already been mentioned, there are other nonradiative relaxation pathways that may be accessible to the excited-state species, such as charge transfer, proton transfer, and Förster resonance energy transfer (FRET).

Förster Resonance Energy Transfer

FRET is a mechanism of energy transfer from an initially excited donor molecule (or moiety) to an acceptor molecule (or moiety). FRET transfer efficiency is defined as the fraction of photons absorbed by the donor that are transferred to the acceptor.[4] As FRET requires dipole-dipole coupling to occur, transfer efficiency is highly dependent on the distance between donor and acceptor as well as the alignment of their transition dipoles. FRET systems are characterised by the Förster distance, R_0, which represents the distance at which energy transfer efficiency is equal to 0.5 (this is typically around 20 to 60 Å).[4] The Förster distance, in Å, can be calculated by[4–6]

$$R_0 = 0.211(J_{DA}\kappa^2\eta^{-4}\Phi_D)^{1/6} \text{ (in Å)}, \qquad 2.2$$

where J_{DA} is the spectral overlap of the donor and acceptor, κ is a geometric factor that accounts for the relative orientation of the donor and acceptor, η is the refractive index of the medium, and Φ_D is the quantum yield of the donor. The quantum yield of a molecule relates to the proportion of photons it emits compared to the number it has absorbed; a more formal definition will be given later. The spectral overlap can be calculated *via*[4,7]

$$J_{DA} = \frac{\int \varphi_D(\lambda) \cdot \varepsilon_A(\lambda) \cdot \lambda^4 d\lambda}{\int \varphi_D(\lambda) d\lambda}, \qquad 2.3$$

where φ_D is the emission spectrum of the donor, ε_A is the extinction coefficient of the acceptor, and λ is wavelength. The orientation parameter, κ, is given by[5,6]

$$\kappa = \boldsymbol{e}_1 \cdot \boldsymbol{e}_2 - 3(\boldsymbol{e}_1 \cdot \boldsymbol{e}_{12})(\boldsymbol{e}_{12} \cdot \boldsymbol{e}_2), \qquad 2.4$$

where $\mathbf{e}_1$ and $\mathbf{e}_2$ are the unit vectors for the donor and acceptor transition dipoles and $\mathbf{e}_{12}$ is the unit vector between their centres. The value of κ^2 can range from 0 to 4 and must be correctly accounted for to accurately determine structural information from experiments. The value of κ^2 is typically set to ⅔ when both the donor and acceptor molecules (transition dipoles) are able to freely rotate within the system (effectively averaging out any orientation dependence). In contrast, when movement within the system is restricted, such as when the fluorophores are incorporated into nucleic acid structure, it is necessary to establish (or account) for the instantaneous orientation of

the fluorophores; this is exemplified by the work of Börjesson *et al.*,[5] Ouellet *et al.*,[7] and Iqbal *et al.*[8]

For a system where the Förster distance is known and fixed, the FRET transfer efficiency, ET_{FRET}, can be calculated *via*[7]

$$ET_{FRET} = \frac{1}{1 + \left(\frac{R_{DA}}{R_0}\right)^6}, \qquad 2.5$$

where R_{DA} is the distance between the donor and acceptor. The sixth power dependence means that the transfer efficiency is extremely sensitive to the distance between the donor and accepter when R_{DA} is close to R_0. This allows the use of FRET to measure very small changes in distance; however, it also means that it is only possible to practically measure distances between $0.5R_0$ and $2R_0$ because the change in efficiency becomes too small to reliably detect outside of this range, as shown in Figure 2.3.

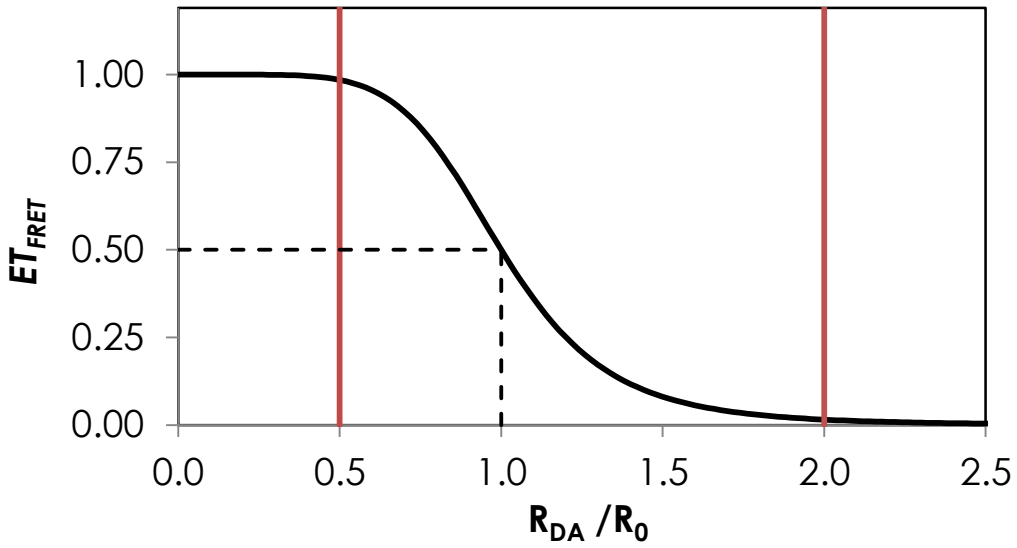

Figure 2.3 – Förster resonance energy transfer efficiency, ET_{FRET}, as a function of the distance between donor and acceptor molecules, R_{DA}, which is given in units of the Förster distance, R_0. The red lines bound the range of distances which are practically measurable using FRET. Figure based on Lakowicz.[4]

It is worth noting that, although it can eventually result in emission from the acceptor molecule, FRET is a nonradiative process from the perspective of the initially excited molecule (or moiety).

2.1.4 Excited-State Dynamics

The stochastic relaxation of a population of excited-state molecules, [*M**], can be modelled by considering both radiative (fluorescence and phosphorescence) and nonradiative (IC, ISC, collisional quenching, and so on) processes. In this case, relaxation from the excited-state is described by first-order rate kinetics;

$$-\frac{d[M^*]}{dt} = (k_R + k_{NR})[M^*] = k[M^*], \qquad 2.6$$

where k_R and k_{NR} are rate constants representing all radiative and nonradiative processes, respectively. As k_R and k_{NR} are indistinguishable during measurements, it is convenient to use the summation $k = k_R + k_{NR}$ for further discussion. Integration of Equation 2.6 yields the expected temporal evolution of the excited-state population;

$$[M^*](t) = [M^*]_0 e^{-kt}, \qquad 2.7$$

where $[M^*]_0$ is the concentration of excited molecules, *M**, at $t = 0$ (typically defined by the arrival of the excitation pulse). The intensity of the observed emission is directly proportional to the excited-state population and so the decay dynamics can be described by

$$I(t) = I_0 e^{-\frac{t}{\tau_f}}, \qquad 2.8$$

where I_0 is the intensity observed immediately after excitation and $\tau_f = 1/k$ is the fluorescence lifetime of the emitting species. The value of τ_f is a particularly useful parameter to use to describe the system under study because it represents the average time interval between excitation and emission; however, it should be noted that only a very small fraction of the initial excited-state population emit at precisely this time. Another important characteristic parameter is the emission quantum yield, Φ, which is given by

$$\Phi = \frac{k_R}{k_R + k_{NR}}. \qquad 2.9$$

The value of Φ represents the ratio of the number of photons emitted to the total number absorbed and has a maximum value of unity when the nonradiative rate constant is zero.

The simple form of Equation 2.8 offers a convenient way to probe the excited-state dynamics of a fluorescent species. By monitoring the attenuation in fluorescence intensity after excitation, it is possible to determine the fluorescence lifetime, which, in turn, can be used to interpret the relaxation dynamics of the excited-state species. One of the most common techniques used to measure the fluorescence decay is time-correlated single photon counting (TCSPC).

2.1.4.1 Time-Correlated Single Photon Counting

A full description of TCSPC will not be presented here but, if further detail is sought, it is comprehensively covered by Lackowicz,[4] O'Connor and Phillips,[9] and Becker.[10] Briefly, the detection process of TCSPC essentially works like a stop-watch. The timer is started when the excitation pulse reaches the sample and is stopped when the first (single) emitted photon arrives at the detector. In this case the arrival of the next excitation pulse resets the timer. Alternatively the timer may be started by the arrival of the emitted photon and stopped by the subsequent excitation pulse; this is a more efficient detection method because it guarantees that timings only occur when an emitted photon is detected. By necessity, the excitation pulse is of considerably shorter duration (fs timescale) than the fluorescence lifetime to be measured (ps to ns timescale). The time interval between the excitation pulse and the detected photon is binned into an accumulating histogram of arrival times, which is representative of the fluorescence decay of the system under investigation. To build up a statistically reliable decay it is necessary to repeat the detection process many (potentially millions) of times. Unfortunately, the probability of detecting a single photon after an excitation pulse has to be kept very low. This ensures that the probability of two photons arriving at the detector after a single excitation pulse is essentially zero. If this were not the case then there would be a bias towards photons of short lifetime because only the first photon is detected (counted); however, this means that acquisition of even a single decay can become time-consuming. The requirement for, among many other things, short excitation pulses and sensitive detection means that there are significant technological challenges to performing TCSPC measurements. In addition, even once successfully measured, analysis and interpretation of the resulting fluorescence decay is not always straightforward; this issue is addressed in more detail in Chapter 4.

2.2 Computational Chemistry

The relentless growth of computing power over the past few decades has led to considerable advances in the application of computational approaches to understand chemical systems. Computational methods offer the opportunity to corroborate (or perhaps discredit) current theories based on experimental data. They may also be able to provide insight into the properties of a system that would otherwise be inaccessible through purely experimental means. This section will very briefly outline some of the key concepts related to computational chemistry. The content has mostly been derived from the work of Jensen,[11] Young,[12] and Gibson.[13]

The main principle of quantum chemical calculations is to obtain the state that has lowest energy; the most stable state. This state can be determined by solving the time-independent Schrödinger equation, which is given by

$$\hat{H}\Psi = E\Psi, \tag{2.10}$$

where $\hat{H}$ is the Hamiltonian operator; Ψ is a wavefunction; and E is energy. Simply stated, the Hamiltonian operator characterises the total energy of the given wavefunction. It is related to the kinetic energy operator, $\hat{T}$, and potential energy operator, $\hat{V}$, by

$$\hat{H} = \hat{T} + \hat{V}. \tag{2.11}$$

The wavefunction is a probabilistic description of electron behaviour. As the name implies, it interprets an electron as a wave; the properties of which are a function of electronic and nuclear positions. The square of the wavefunction yields a probability function that allows calculation of the *probability* of an electron being at a particular location in space (and time). The time-independent Schrödinger equation is an example of an *eigenvalue equation*; the wavefunction is an eigenfunction of the Hamiltonian that corresponds to eigenvalue E, which is interpreted as energy. Computational approaches that attempt to determine the wavefunction by solving the Schrödinger equation are known as *ab initio* methods since they arise from the first principles of quantum mechanics.

The succinct form of Equation 2.10 is deceptive and exact solutions to the Schrödinger equation have only been found for simple, one-electron systems, such as the hydrogen atom. To be able to study more complex systems it is necessary to introduce approximations. As has been outlined previously, the Born-Oppenheimer approximation allows the separation of the electronic and nuclear parts of the wavefunction by assuming that the nuclei are static. The electronic Schrödinger equation with *clamped nuclei* is of the form

$$\hat{H}_{el}\Psi_{el}(r\,;R) = E(R)\Psi_{el}(r\,;R), \qquad 2.12$$

where the notation $(r\,;R)$ represents the fact that although the wavefunction is a function of the electron coordinates, r, it also has parametric dependence on the nuclear coordinates, R. In other words, the solution of Equation 2.12 is specific to the nuclear coordinates (the geometry) of the molecule. It is therefore evident that the optimal molecular conformation of the system could be determined by iteratively solving Equation 2.12 for different sets of nuclear coordinate until a minimum in energy is achieved. Unfortunately, the fundamental issue of solving the Schrödinger equation has still not been resolved. The electronic Hamiltonian operator is given by

$$\hat{H}_{el} = \hat{T}_e + \hat{V}_{en} + \hat{V}_{ee} + \hat{V}_{nn}, \qquad 2.13$$

where $\hat{T}_e$ describes the kinetic energy of the system of electrons and $\hat{V}_{en}$ describes the Coulombic attraction between electron and nuclei. The term $\hat{V}_{nn}$ is related to nuclear-nuclear repulsion and is typically accounted for at the end of the calculation since it depends only on stationary nuclei. The term $\hat{V}_{ee}$ represents electron-electron interactions and turns out to be the main source of difficulty in solving the Schrödinger equation for a many-electron system. The problem stems from the fact that the motion of each electron is *correlated* to all other electrons in the system. There are two main effects related to electron correlation; exchange and Coulomb correlation. Exchange correlation is a manifestation of the Pauli exclusion principle. Since the wavefunction is required to be anti-symmetric with respect to exchange of electrons, it is necessary for electrons with parallel spin to be separated in space; this leads to an effective repulsion between the electrons. Coulomb correlation is a direct consequence of the electrostatic repulsion of the negatively charged electrons. Computational methods must address

correlations effects to be able to solve the Schrödinger equation for a many-electron system.

The simplest type of *ab initio* calculation follows the Hartree-Fock (HF) method. HF calculations are founded on the uniform field approximation, which means each electron experiences the average repulsion potential of all other electrons rather than directly interacting with each individual electron. This approach has the advantage of simplifying a many-electron wavefunction into a combination of simpler, one-electron wavefunctions, which are known as spin-orbitals. The HF method accounts for exchange correlation through the use of Slater determinants. Rather than use a simple (Hartree) product of spin-orbitals to describe the overall wavefunction, Slater determinants provide linear combinations of spin-orbitals that inherently respect the requirement for anti-symmetric exchange. The main disadvantage of using the uniform field approximation is that Coulomb correlation is neglected, which can lead to large deviations from experimental results.

Two common approaches that attempt to address the electron correlation deficiency of the HF method are configuration interaction (CI) and Møller-Plesset (MP) perturbation theory. The CI method allows the electrons to occupy virtual orbitals. This provides more space for the electrons to avoid each other and, in turn, ensures a better description of their correlated motion. The MP approach attempts to account for the correlation effects missing from the HF method by introducing a perturbation, $\hat{V}$, into the Hamiltonian operator, such that;

$$\hat{H} = \hat{H}_0 + \lambda \hat{V}, \tag{2.14}$$

where $\hat{H}_0$ is the unperturbed Hamiltonian operator, which is obtained from the HF method, and λ is an arbitrary parameter that controls the size of the perturbation. The perturbed wavefunction and perturbed energy are then expressed as a power series in λ. The level of theory that is used for calculations is denoted MP*n*, where *n* is the order of the perturbation (the highest power of λ used). CI and MP methods can significantly improve the accuracy of calculations but the trade-off is that they require considerable computational effort. In practice, it is only possible use these methods for small systems.

An alternative to *ab initio* methods is density functional theory (DFT). Rather than attempt to solve the Schrödinger equation, the aim of DFT is to optimise with

respect to the electron density. The basis for this approach comes from an early proof by Hohenberg and Kohn[14] that showed the ground-state energy is directly related to the electron density of the system. Kohn and Sham[15] further developed DFT by introducing the concept of using orbitals, similar to the strategy of *ab initio* methods. DFT has become popular because of it can replicate the results of methods with similar accuracy at much lower computational cost. The use of the term *functional* denotes the fact that the calculated energy is a function of the electron density, $\rho(r)$, which is itself a function of the Kohn-Sham orbitals, φ_α. For a system of N electrons, the density is given by

$$\rho(r) = \sum_{\alpha=1}^{N} |\varphi_\alpha(r)|^2. \qquad 2.15$$

Unfortunately, the exact functional that should be used to calculate the energy from the obtained density is not known. This has led to number of different proposed functionals that each have their own advantages and disadvantages; the M06-2X functional is used in this work.[16–18] M06-2X supplements underlying exchange and correlation functionals, which directly calculate the electronic energy, with a large number of energetic parameters derived from the thermochemical data of dispersion-bound systems. Compared to typical DFT methods, M06-2X has greatly improved performance for systems where dispersion is important. M06-2X is an example of a hybrid functional because it incorporates a portion of the exact exchange obtained from HF theory.

Another necessary consideration for computational calculations is the basis set that is to be used. Basis sets aim to reproduce the underlying orbital structure of the molecular system. Orbitals are generally approximated by the following functional form

$$\varphi_\alpha = Y_{lm} \sum_i C_i \sum_j C_{ij} e^{-\zeta_{ij} r^2}, \qquad 2.16$$

where the spherical harmonic function, Y_{lm}, relates to the orbital symmetry (*s*, *p*, *d*, and so on); $e^{-\zeta_{ij} r^2}$ are Gaussian *primitive* functions; C_i are orbital coefficients; and *contraction* coefficients, C_{ij}, and exponents, ζ_{ij}, are constants that define the basis set. This formalism allows calculations to be performed efficiently because only the orbital coefficients require optimisation. Larger basis set provide a better representation of

the atomic orbitals but incur greater computational cost during calculation. There is therefore a trade-off between accuracy and calculation time although, in some cases, smaller basis sets can give results closer to experimental values because of serendipitous cancellation of errors. Slater-type orbitals (STOs, $e^{-\zeta_{ij}r}$) are an alternative form of primitive function to the Gaussian-type orbitals (GTOs) shown in Equation 2.16. STO basis sets require fewer primitive functions to describe the orbital accurately; however, GTO basis sets benefit from having well-defined analytical properties, which leads to significant computational savings during calculations. In this work, the 6-31+G(d) basis set is used, which follows the *split-valence* notation derived by the Pople Group.[19] Each core orbital is described by a single contraction of six GTO primitives (i = 1, j = 6) while each valence shell orbital is described by two contractions, one of three primitives and the other of one primitive (i = 2, j = 3 or 1). The single plus sign denotes the fact that all atoms except hydrogen have been *augmented* by the addition of diffuse functions (a double plus would indicate the inclusion of hydrogen). Diffuse functions aid in the description of long range interactions, such as van der Waals interactions, which are important for the systems studied in this work. The 'd' in parenthesis means that a set of *d*-type polarisation functions has been added to all atoms (except hydrogen). Polarisation functions can improve the accuracy of geometry calculations and computed vibrational frequencies. The 'G' simply denotes the fact that it is a GTO basis set. The combination of the functional M06-2X with basis set 6-31+G(d), which is succinctly denoted as M06-2X/6-31+G(d), was found have comparable performance to high-level calculations for dispersion-dominated systems, including nucleobase dimers.[20]

The influence of solvent is another factor that has to be considered for computational calculations. Explicit models for solvation are most accurate but can quickly become prohibitively expensive due to the number of solvent molecules that may be required. Implicit models, such as the polarizable continuum model (PCM),[21] offer a compromise by simulating a reaction field which roughly mimics the presence of solvent. In the PCM, solvent interaction terms are calculated using point charges located on the surface of the solute cavity. The cavity is defined by a set of interlocking van der Waals-spheres centred at the atomic positions of the solute molecule.

Once the computational approach that is to be used has been established, the aim of the calculations is to determine the set of orbital coefficients that provides the optimal (lowest) energy for the system. Unfortunately, as mentioned above, electron correlation means that analytical solutions of many-electron problems are unknown.

Electron correlation also means that the system is dependent on itself. In other words, optimising the energy of the system requires the determination of unknown orbital coefficients that are dependent on their own values. Numerical methods are required to solve this type of problem and typically an iterative approach known as the *self-consistent field* (SCF) method is used. This approach works in the following manner. An initial guess is made at the orbital coefficients and the resulting system is 'solved' to obtain a new set of orbital coefficients. In turn, the new set of orbital coefficients defines a new system that can be solved to obtain a newer set of orbital coefficients. Each time the system is solved the set of orbital coefficients is refined. This process continues until (hopefully) convergence is achieved and the resulting set of orbital coefficients is self-consistent; this means that, within the chosen convergence limits, the system defined by the new set of orbital coefficients is unchanged to that initially defined. The set of orbitals defining the self-consistent field can then be used to calculate the properties of the system.

2.3 DNA Structure

The following section will outline the basic structural features of deoxyribonucleic acid (DNA). Considerable effort has been made to use standard definitions for nucleic acids[22] when discussing molecular structure but there are some instances where unusual or uncommon terminology will be used.

2.3.1 Fundamentals of DNA Structure

An example of the well-known, but nonetheless remarkable, double-helical structure of biopolymer DNA is illustrated in Figure 2.4. Colours for atoms are as follows; carbon, brown; oxygen, red; nitrogen, blue; phosphorous, orange; and hydrogen, white. Although there are other variations of nucleic acid structure (such as ribonucleic acid (RNA) and other DNA conformations, which are discussed in §2.3.2) it should be assumed that, unless otherwise stated, discussion within this thesis relates to the B-form of DNA.

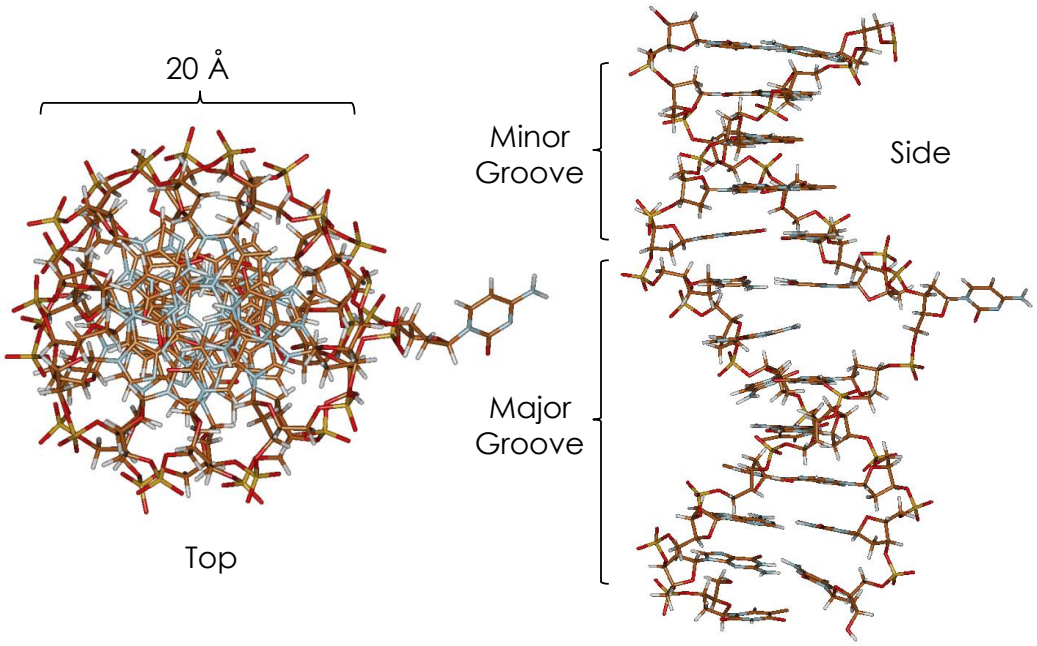

Figure 2.4 – DNA double helix taken from an X-ray crystal structure of a complex containing base-flipping enzyme M. HhaI (PDB 3MHT).[23] The labels 'top' and 'side' are arbitrary but are given to show relative orientation. The duplex shows a flipped base which is held in an enzymatic pocket (for clarity the enzyme has been omitted).

The two entwined anti-parallel strands of DNA are made up of nucleotide units which link together to create extended structures. Each nucleotide is constructed of a deoxyribose sugar group, a phosphate group, and a nitrogenous nucleobase. As a brief note on terminology, the term *base* will be used interchangeably with the term *nucleobase*; however, the latter term will generally only be used where additional clarity might be required. Likewise, the terms *duplex* and *double helix* (referring to the type of structure shown in Figure 2.4) should be considered synonymous. The sugar and phosphate groups in DNA act as a backbone to the polymeric structure while the base (or rather, sequence of bases) encodes the detail of the genetic instruction held by the DNA. Figure 2.5 shows the general structure of the sugar-phosphate backbone of DNA.

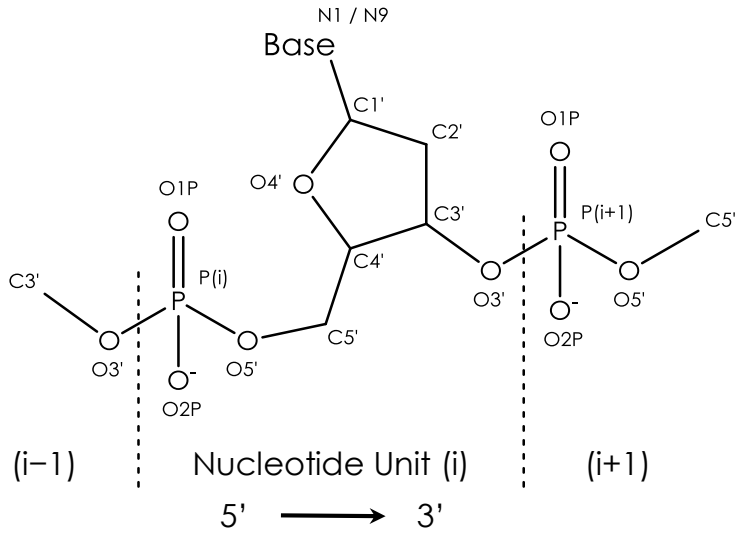

Figure 2.5 – Polynucleotide structure with standard numbering. Units are connected *via* a sugar-phosphate backbone. Each nucleobase (base) is attached through an N-glycosidic linkage. Convention requires that DNA sequences are presented in a 5' to 3' direction. The nomenclature 5' and 3' relates to the numbering of the sugar carbon atoms. Primes indicate that the atoms are part of the backbone (rather than the base).

The direction of DNA is defined by the atomic numbering of the sugar group and sequences are conventionally presented in a 5' to 3' direction. Primes designate atoms that are part of the backbone rather than a base. All nucleotides share the same general structure and only differ by the base which is attached. There are four main DNA bases; adenine (A), cytosine (C), guanine (G), and thymine (T). Their structures are either purine-based (A and G) or pyrimidine-based (C and T) as shown in Figure 2.6. Noncanonical bases 2-aminopurine (2AP) and hypoxanthine (I) are also relevant to the work of this thesis. 2AP is a fluorescent base analogue that will be discussed in more detail below (§2.3.3.1) and I is an analogue of guanine that is sometimes used in studies because of its redox properties. Of particular importance to this thesis is the fact that I is essentially inactive to electron transfer with excited-state 2AP (2AP*).[24,25] It is worth mentioning here that RNA differs from DNA in the fact that the base uracil (U) is present instead of thymine (the only structural difference between these two bases is a methyl group at the C5-position). As the name suggests (and shown later in Figure 2.8), RNA also differs from DNA in that the sugar group is ribose rather than deoxyribose.

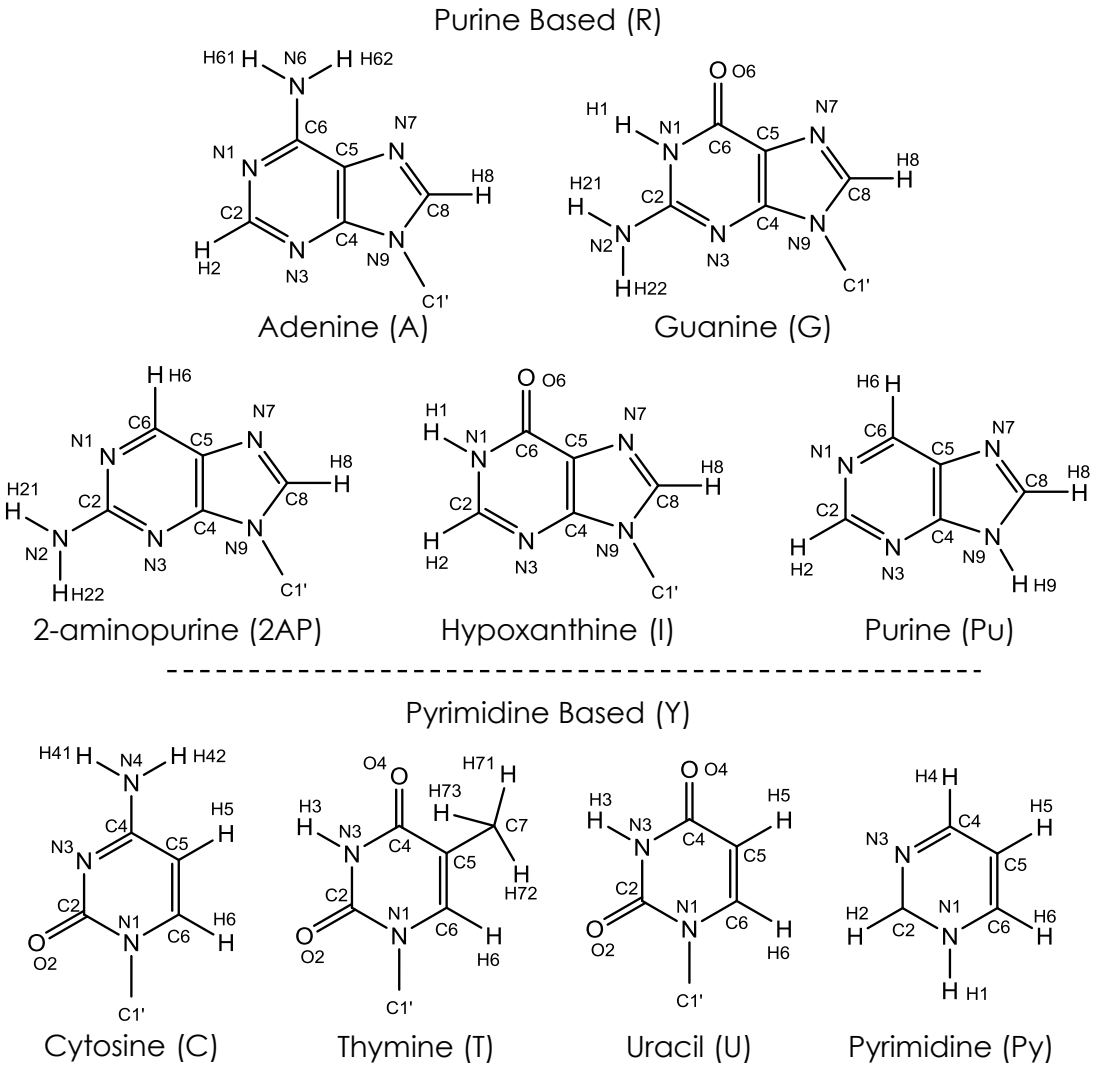

Figure 2.6 – Standard atomic numbering[22] for nucleic acid bases. The bases can be split into two groups (purine or pyrimidine) based on their underlying skeleton structure. In nucleic acids the bases are linked to the sugar-phosphate backbone by an N-glycosidic bond between N1 (pyrimidine) or N9 (purine) and C1' atoms (see Figure 2.5). When the base is in isolation the C1' atom is replaced by a hydrogen atom; as shown in purine and pyrimidine structures.

The bases have various different properties due to their functional groups. For instance, Figure 2.7 shows the differences observed in the total dipole moment of the bases. The orientation and magnitude of the dipole moment is given by the direction and length, respectively, of the black arrow. Comparing adenine and 2-aminopurine shows that moving the NH_2 group from the C6-position (A) to the C2-position (2AP) has a dramatic effect on the direction of the dipole moment. The reduction in the

magnitude of the dipole moment of hypoxanthine compared to guanine shows the effect of removing the NH_2 group at the C2-position.

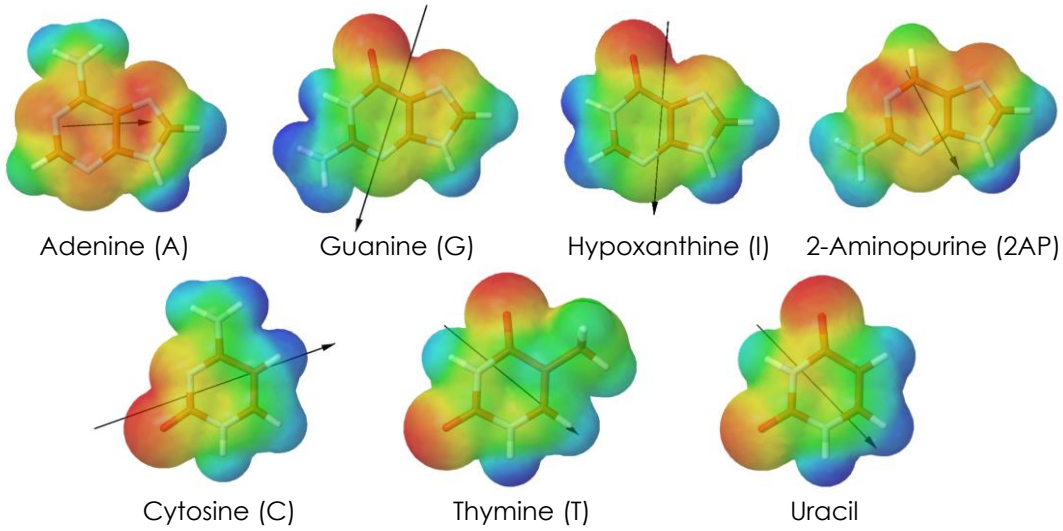

Figure 2.7 – Natural DNA bases (adenine, cytosine, guanine, and thymine) with dipole moments (black arrows) and electrostatic potential mapped onto the solvation shell. The colour scale runs from blue, representing positive charge, to red, representing negative charge. RNA base uracil and noncanonical bases 2-aminopurine and hypoxanthine are also shown.

In addition to representing nucleobases, the notation A, C, G, T, U, I, and 2AP will also be used to represent associated nucleoside and nucleotide structures. Table 2.1 gives a summary of the possible meaning of the codes. The context in which the term is used should unambiguously inform the reader which type of molecular structure is being discussed. It is worth highlighting that the nucleoside associated with the noncanonical base hypoxanthine is known as inosine, hence the use of symbol I (rather than H).

Code	Nucleobase	Nucleoside	Nucleotide	
A	Adenine	Adenosine	Adenosine monophosphate	Adenosine triphosphate
C	Cytosine	Cytidine	Cytidine monophosphate	Cytidine triphosphate
G	Guanine	Guanosine	Guanosine monophosphate	Guanosine triphosphate
T	Thymine	Thymidine	Thymidine monophosphate	Thymidine triphosphate
U	Uracil	Uridine	Uridine monophosphate	Uridine triphosphate
I	Hypoxanthine	Inosine	Inosine monophosphate	Inosine triphosphate
2AP	2-aminopurine	2-aminopurine nucleoside	2-aminopurine nucleoside monophosphate	2-aminopurine nucleoside triphosphate

Table 2.1 – Terminology associated with the structural components of nucleic acid constituents. The four main bases in DNA are adenine, cytosine, guanine, and thymine. In the case of RNA, uracil is present instead of thymine. Noncanonical bases 2-aminopurine and hypoxanthine are base analogues relevant to work in this work.

The structural difference between nucleobases, nucleosides, and nucleotides are shown in Figure 2.8.

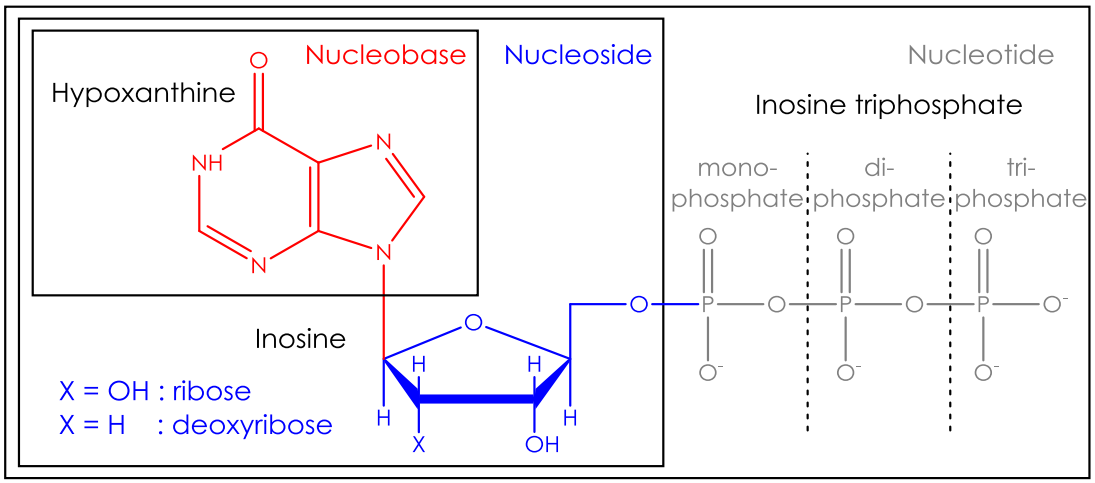

Figure 2.8 – Structural components of nucleic acid constituents. Colours represent distinct structural sections that form a nucleobase (red only), a nucleoside (red and blue), or a nucleotide (red, blue, and grey). The labels in black text relate to the specific molecule bounded by each box.

As almost the entire content of the thesis relates to DNA (deoxyribonucleic acid) rather than RNA (ribonucleic acid) is should be assumed that the *deoxy* form (without the C2' hydroxyl group) of the sugar is present within the molecular structure being discussed. For instance, when talking about adenosine it should be assumed this refers to deoxyadenosine. In situations where it is required, the usual notation of 'r' and 'd' will be used to distinguish the sugar type present; for example, dC would be deoxycytidine while rC would be cytidine.

In the context of nucleic acid structure, the symbol N will be used where aNy natural base (A, C, G, or T) is present. In Chapter 3 this terminology is extend to also include I (that is, N = A, C, G, T, or I). The symbol R will be used to represent a puRine based nucleobase (A or G). The symbol Y will be used to represent a pYrimidine based nucleobase (C or T). The notation 'Pu' and 'Py' will be used to donate purine and pyrimidine molecules, respectively (note this is not the same as R and Y). Where it is required, the usual notation of NMP or NTP will denote nucleoside monophosphates and nucleoside triphosphates, respectively.

A significant proportion of the work contained within this thesis relates to a molecular structure that has been called a 'dinucleotide'; Figure 2.9 shows a schematic of the general structure of this type of molecule.

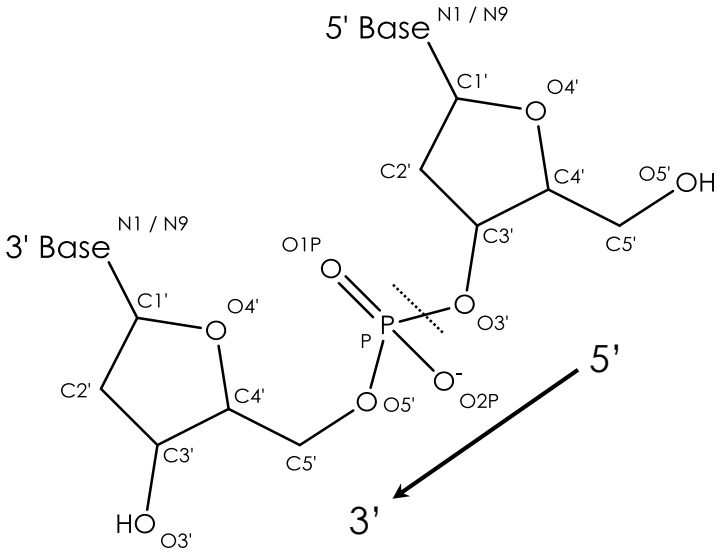

Figure 2.9 – Schematic of the molecular structure of a dinucleotide; a fragment of DNA that contains two bases that are connected together with a sugar-phosphate backbone. Directionality is defined by the C5' and C3' atoms of the sugar group. The dotted line between O3' and P atoms highlights the crossover from the 5' nucleoside (which does not contain a phosphate moiety) to the 3' nucleotide.

It is worth pointing out that the use of the term dinucleotide is a slight misnomer; this is because the 5' unit does not contain a phosphate group and is therefore an example of nucleoside. A more correct description of the molecule in Figure 2.9 would be dinucleoside monophosphate; however, for brevity the term dinucleotide will be used throughout this thesis. In the studies conducted the 5' base was either 2AP or A while the 3' base was any of A, C, G, T, or I.

As convention requires, consecutive bases in a sequence of bases are always presented in a 5' to 3' direction. Dinucleotides (of the form shown in Figure 2.9) will be represented with a hyphen between base codes. For instance, 2AP-T is a dinucleotide with 2AP as the 5' base and T as the 3' base. Nucleobase dimers, which are missing the sugar-phosphate backbone and are not covalently bonded together, will be represented with a vertical line between base codes. For instance, 2AP|T is the notation for a dimer with 2AP as the 5' base and T as the 3' base. Although the terms 5' and 3' are effectively meaningless without a phosphate backbone present, the base ordering gives the sequence of the parent dinucleotide structure that was used to create the dimer. The difference between 2AP-T and 2AP|T can be seen by the comparison in Figure 2.10.

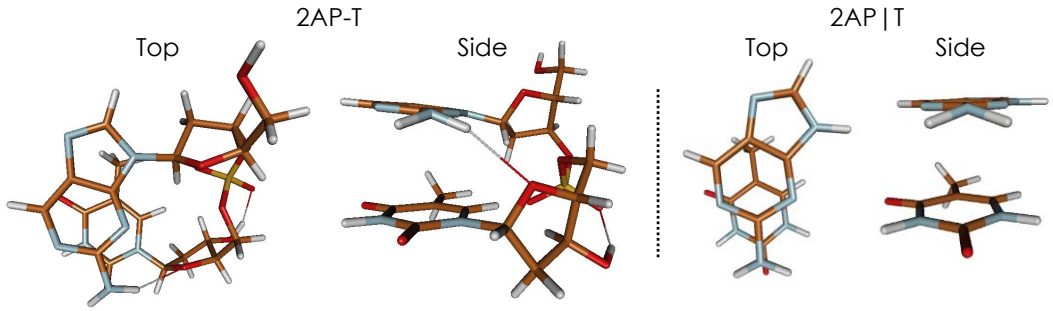

Figure 2.10 – Comparison of the molecular structure of dinucleotide 2AP-T and dimer 2AP|T. The labels 'top' and 'side' are arbitrary but are given to show relative orientation.

The helical shape of double-stranded DNA is caused by stacking interactions between neighbouring bases of the same strand (base-stacking) and hydrogen bonding (H-bonding) between pairs on opposite strands (base-pairing). Bases lie almost parallel to their flanking neighbours with a translational (rise) and rotational (twist) offset of around 3.4 Å and 36°, respectively. The arrangement of bases induced by base-stacking provides a convenient geometry for base-pairing to occur. In this thesis, base pairs will be represented with a middle dot between the base codes. For instance, G·C represents a base pair between bases G and C. In a duplex, complementary bases form H-bonded

pairs between opposite strands; the most favourable parings are those predicted by Watson-Crick, namely A·T and G·C, and are shown in Figure 2.11.

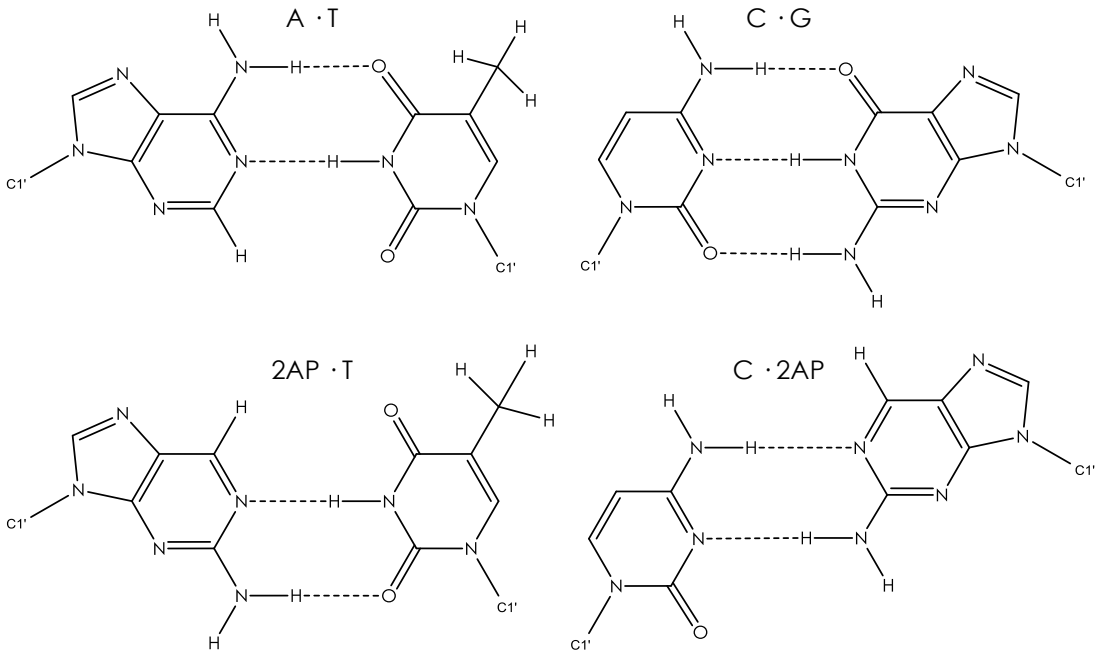

Figure 2.11 – Watson-Crick base-pairing in natural bases A, C, G, and T. Dashed lines represent H-bonds between bases. Base analogue 2AP can form a Watson-Crick base pair with T and a wobble base pair with C.

The G·C pairing contains three H-bonds while the A·T pair only has two and is inherently less stable. Other geometries such as Hoogsteen base-pairing and wobble base-pairing are also possible. These alternative conformations are not as stable in the double-helical form of DNA but do allow the formation of alternative structures such as a triple helix (triplex). The length of a duplex is typically defined by the number of base pairs (bp) it contains. DNA sequences can reach millions of base pairs in length.[26,27]

The double helix of DNA is, without doubt, one of the most iconic structures to be found in nature. It is therefore difficult to contemplate that just over 60 years ago its structure was still merely speculation. Clearly the intricacies of the underlying structure have taken a great deal of effort to resolve and there are many people that are responsible for the knowledge that exists today. The story of DNA is a fascinating one and it is worthwhile knowing where the current description has come from. An excellent overview of the history surrounding DNA structure is provided by Berman and Olson[28] but a brief outline will be presented below. This will lead on to developments that have been made more recently and the techniques that are now providing even greater insight into DNA structure.

2.3.2 A Brief History of DNA Structure

Although Meischer first isolated '*nuclein*' in 1869, it took until the early 1950s for the significance of DNA to biological function to be established.[29] Experiments by Griffith had shown that nonvirulent bacteria could be made virulent in a process later known as transformation (where genetic alteration occurs).[30] This provided evidence that genetic material (of some kind) could be transferred between bacterial cells. Chemical analysis (along with some enzymatic experiments) of the same system by Avery, MacLeod, and McCarty suggested that DNA was responsible for genetic transformation.[31] At the time there was also a competing hypothesis that proposed genetic information was held by proteins; however, the work of Hershey and Chase on the T2 bacteriophage (a virus) dismissed this postulate in the favour of DNA.[32] Their experiments showed that DNA is injected into the bacterial cell while the viral protein capsule remains mostly outside. Thus, genetic changes within the cell were concluded to have been caused by DNA.

A race was then on to determine the structure of this important biological molecule. The composition of nucleic acids had been determined in earlier work by Kossel[33] and Levene[34] and there was evidence that DNA was built as a string of nucleotide units (the structure of nucleotides will be discussed in detail in the next section). The X-ray crystal data collected by Astbury gave weight to the idea of a regular helical structure but were not of high enough quality to discriminate between different models.[35] For example, Pauling and Corey suggested a triplex structure in which nucleobases radiated outwards from a polymeric sugar-phosphate core.[36] While this proposal turned out to be wrong, a successful model of DNA was coming ever closer to fruition. A paper by Furberg[37] exemplifies the iterative nature of the process at the time; at each attempt previous ideas were being built upon and, in turn, provided their own piece of the picture.

In 1953 Watson and Crick proposed their model of the DNA double helix.[38] Their structure was confirmed (and enlightened) by the superior X-ray crystal data of Wilkins[39] and Franklin.[40,41] Perhaps the most important aspect of the Watson-Crick model was base-pairing.[42] This critical insight had been greatly influenced by earlier work of Chargaff *et al.*[43,44] and Wyatt[45] which showed that ratios of adenine to thymine and guanine to cytosine were close to unity. Even at this early stage the implications base-pairing had for genetic replication were realised. In fact, unbeknown to many at

the time, the idea of complementary strands holding the key to heredity had already been postulated decades before by Koltsov.[46]

Further work by Franklin and Gosling provided evidence that the Watson-Crick double helix was not the full story though.[47] Distinct X-ray diffraction patterns, labelled simply A and B, could be obtained depending on water content and preparation history of the DNA crystal. These conformations have since become known as A-DNA and B-DNA with (in a slight twist of fate) the Watson-Crick duplex being most closely related to the B-form.[a] Subsequent investigation found many other DNA conformations were possible but the B-form dominates in the conditions found in the cell.[48–50]

Despite the award of the Nobel Prize in Physiology or Medicine in 1962 to Watson, Crick, and Wilkins there was still some lingering doubt about the accuracy of the proposed DNA duplex structure. Although the general conformation of DNA had been established, the resolution of the X-ray data was still too poor to provide detail at the atomic level. It was therefore difficult to validate the idea of specific pairing between particular bases. Small fragments of DNA were investigated in an attempt to verify the conformational relationship between the bases. These simplified systems allowed greater insight into specific base-pairing interactions that occurred. As the simplest possible model of DNA, substituted bases were initially co-crystallised to look at H-bonding patterns.[51,52] Substitution was used to try and ensure bases paired the way they were thought to be orientated in DNA. Interestingly the obtained structure was in an alternative conformation than the base-pairing predicted by Watson and Crick. In essence, the purine-type base was over-turned from its expected geometry and a different edge was used for base-pairing. This type of conformation has become known as Hoogsteen base-pairing in respect of the experimentalist who discovered these structures.

After the work of Hoogsteen, a succession of X-ray crystallography[53–63] and NMR[64–68] experiments were performed on dinucleotides. Most of these early experiments focussed on RNA constructs, presumably because they were easier samples to obtain than equivalent DNA dinucleotides.

One particular sequence that was investigated in many studies was uridylyl-3',5'-adenosine phosphate (UA). A review by Kim *et al.* of the experimental data collected for this dinucleotide suggested that there were 7 distinct conformations

[a] The Franklin and Gosling paper was submitted before the famous Nature triplet was published.

possible for nucleic acids.[69] A more recent paper proposes that, in fact, as many as 32 distinct dinucleotide[b] conformations are possible in RNA structures.[70]

The work on dinucleotides highlighted the fact that crystal packing might have an influence on the observed nucleic acid structures. Unlike globular proteins, DNA constructs were found to crystallise without a solvation shell.[28] Intermolecular interactions between neighbouring duplexes could therefore potentially cause structural perturbation.

Although there was a possibility that crystal packing caused structural distortion these studies proved to be very important in the understanding of DNA conformation. The dinucleotide structures confirmed that Watson-Crick base-pairing was a favourable geometry.[58–61] This was a vital discovery as it validated the principal element of the Watson-Crick model; specifically, that the bases paired in the following way; adenine with thymine (uracil) and guanine with cytosine.

The concept of base-pairing was also bolstered by a significant number of studies which utilised infrared (IR) spectroscopy. A succinct overview of the early development of the IR methods used to study DNA is given by Barth and Haris.[71] A particularly noteworthy advance was the transition from solid-state[72,73] to aqueous media.[74] The work of Falk *et al.*[75] and Thomas[76] was also essential for obtaining a proper interpretation of the IR spectra collected.

Although not discussed in detail here, it is worth noting that further developments have made IR spectroscopy an increasingly valuable tool for the study of DNA. For example, Elsaesser gives a summary of recent time-resolved experiments that were used to investigate the vibrational dynamics of base pairs.[77] Excited state dynamics in DNA have also been probed using time-resolved IR spectroscopy.[78–81] As another example showing the versatility of IR techniques, Fourier Transform IR (FTIR) has recently been used to investigate counter-ion binding in the major and minor grooves of DNA.[82]

Despite the apparent success of the Watson-Crick model, alternative structures, such as the very aesthetically displeasing side-by-side model proposed by Rodley *et al.*,[83] were still being postulated as late as the 1970s. One of the main challenges to overcome at the time was the lack of suitable samples to study. Fortunately developments made in the synthesis of oligonucleotides helped to alleviate this issue.[84]

---

[b] Here *dinucleotide* simply means two consecutive bases (potentially within a longer construct).

This approach was hardly ideal though as it was incredibly slow. For example, a 21-mer DNA duplex took the equivalent of 4 years' worth of effort to produce.[85]

Although synthesising DNA was laborious it opened up the opportunity to study novel constructs that had much longer sequences than a dinucleotide. Extension of the structures was important to obtain a better comparison with the proposed models. The dinucleotides might have conformed to base-pairing but any conclusions made about the tertiary structure (such as the helical repeat[c]) of DNA were just speculation through extrapolation.

In 1979 the first DNA structure with atomic resolution appeared.[86] Despite the knowledge and prediction of many different possible double-helical conformations the structure that was obtained was somewhat of a surprise. It was coined Z-DNA because of the zigzag pattern imposed on the sugar-phosphate backbone by having a conformational repeat involving two base pairs (rather than one).[87] It also differed from the Watson-Crick double helix in its handedness (being left- rather than right-handed) and by the fact the bases were upside-down from their typical orientation. The short and repetitive $d(CG)_3$ sequence was likely responsible for this relatively rare conformation being obtained but confirmation of B-DNA was tantalisingly close.[88] The first high-resolution structure of B-DNA was determined in 1981.[89] The construct, now commonly referred to as Dickerson-Drew DNA, had a sequence of AATT between the CG repeat. This addition was evidently enough to stop formation of the Z conformation in favour of the B-form. The atomic structure of A-DNA was soon to follow.[90] The A-form has many similarities to B-DNA but is a slightly more compact structure; the asymmetry of minor and major grooves found in B-DNA is diminished and the duplex has a greater circumference. In addition, while the bases of B-DNA are almost perpendicular to the helical axis they have a much greater inclination in A-DNA.

Although X-ray crystallography and NMR spectroscopy were the main sources of structural information at the time, alternative methods were also used to gain some insight into DNA conformation. An elegant approach developed by Wang utilised electrophoresis to separate topological isomers of DNA in solution.[91] This allowed the helical repeat to be accurately determined through band-shift measurements. Incidentally, the results of this study also dismissed the Rodley *et al.* side-by-side model of DNA.

---

[c] The number of base pairs in a (360°) turn of the DNA helix.

A collection of studies employing the electrophoretic band-shift technique was subsequently investigated by Kabsch, Sander, and Trifonov.[92] Their approach was to model the sequence of each helical repeat into a linear system of equations. By solving this set of equations they were able to extract sequence-specific twist angles. Their results showed that there was significant sequence-dependence to the twist angle between bases.

The sequence-dependence of the twist angle is indicative of the heterogeneous nature of the DNA double helix. It is convenient (and often very useful) to think of a duplex in an idealised form but the reality is far more complicated. Gaining insight into the cause of the sometimes subtle differences that can be present throughout the duplex requires an in-depth and rigorous evaluation of a very large parameter-space. While there have been major improvements in DNA synthesis techniques in recent years there are many challenges to address when preparing samples for crystallographic studies.[93] The difficulty of obtaining suitable crystals is perhaps one of the reasons why that, even today, there are relatively few DNA structures available for evaluation. Currently in the Nucleic Acid Database (NDB)[94,95] there are a total of 7,095 structures (from both X-ray crystallography and NMR experiments). Of those 4,654 are related to DNA but for structures that contain only DNA there are just 1,503. This number is further reduced to 611 when B-DNA is explicitly searched for and many of these structures contain modifications (such as abasic sites or base mismatches) from canonical structure. Additionally, the sequences of DNA are often quite short (fewer than 20 base pairs) and so were perhaps more susceptible to anomalous effects at strand termini. Structures that contain 2AP (the need for which will be discussed later) are even scarcer with only 18 in total. The solution NMR study by Dallmann *et al.* provides the sole 2AP-containing, duplex-only structure.[96] As a rough comparison of numbers, the Protein Data Bank (PDB)[97–99] contains a total of 98,720 structures and is now increasing in size by almost 10,000 structures a year.[d]

In addition to the difficulty of obtaining suitable samples, X-ray crystallography is limited by the fact it offers only a snapshot in time. The structure of DNA is known to fluctuate considerably in solution and this adds even greater complexity to the system. New approaches have therefore been required to gain further insight into the structural properties of nucleic acid constructs.

[d] NDB and PDB values correct as of 20/03/2014

Early circular dichroism (CD) experiments provided useful insight into the conformation of short olignucleotides in solution.[100–102] More recently, CD has been used to characterise the perturbation caused by the presence of base analogues in longer nucleic acids.[103–105]

Ion-mobility mass spectrometry measurements performed by Gidden and Bowers provided insight into the dinucleotide conformations which are both accessible and stable in the gas-phase.[106] Molecular dynamics simulations were used to confirm structurally distinct families which had been identified by differences in observed cross-sectional areas. The results suggested that both open and stacked dinucleotides conformations are present in the gas-phase. As will be seen, this is a particularly pertinent result because it gives weight to the idea that dinucleotides can mimic the conformational states thought to exist within a duplex construct.

Computational methods have become increasingly valuable over the past few decades due to their improved accuracy, reliability, speed, versatility and, to some extent, ease of use. A whole suite of tools is now available to investigate almost any conceivable system. The applicability of computational methods to nucleic acid structures has not been lost on the scientific community and a great body of work has accumulated in recent years. Chapter 5 will expand on the efforts that have been made to gain insight into the conformational dynamics of DNA by using computational approaches.

Due to their sensitivity, flexibility, and information-rich results, fluorescence based techniques offer great potential for providing new insights into the complex dynamics of DNA. Unfortunately, as the natural nucleic bases are essentially non-fluorescent, it is not possible to study purely natural DNA system with common fluorescence methods; however, this issue can be circumvented by incorporating a suitable emissive reporter. The following section will present a brief overview of the use of fluorescent base analogues for the study of nucleic acid systems.

2.3.3 Fluorescent Nucleic Acid Base Analogues

This section will give a brief overview of the use of fluorescent base analogues for the study of nucleic acid systems; a more complete discussion can be found elsewhere, such as in the reviews by Millar,[107] Wilson and Kool,[108] or Sinkeldam, Greco, and Tor.[109] Wilhelmsson also gives a succinct summary of a number of different fluorescent base analogues that have been developed in recent years.[6] A particularly useful idea from this review is the distinction between external and internal modification. Integrating a

fluorophore (modifier) outside of the base-stacked structure of the nucleic acid can be considered to be external modification. In this case the fluorophore is covalently attached to the nucleic acid by a molecular linker. Attachment might occur directly to the sugar-phosphate backbone or could instead occur with a base moiety in the oligonucleotide sequence. The cyanine labelled cytidine nucleotide used in Chapter 6 is an example of an external modification. Internal modifiers are base analogues that are incorporated directly into the base-stacked structure of the nucleic acid and mimic the interactions of a natural base. The most commonly used internal modifier for nucleic acids is 2-aminopurine (which will be discussed in detail below because of its relevance to this thesis) but there are many other examples that exist; indeed, the review by Wilhelmsson is almost entirely concerned with internal modifiers.[6] Some of the advantages and disadvantages of the two types of modifier will now be discussed.

External modifiers can focus more on the photophysical aspects of the problem at hand rather than the physiological ones. The fluorophore can therefore be designed to have very high fluorescence quantum yield (Φ_f) and extinction coefficient (ε) and, consequently, means they can be extremely bright (brightness $\propto \varepsilon \cdot \Phi_f$). Improving brightness gives a greater signal to detect and so reduces the detrimental effects of noise. Focussing on photophysical properties also provides opportunity to improve the photostability of the fluorophore. Resistant to possible degradation cause by incident radiation is crucial if the fluorophore is to be useful as a reporter. Greater freedom in the design of the fluorophore also offers the prospect of being able to tune the excitation and emission wavelength of the reporter to suit the task at hand. Moving from ultraviolet (UV, where the natural bases absorb) into the visible or IR spectral region has a number of advantages including the potential for selective excitation. Lowering the excitation energy will also likely limit photostability issues (for both the reporter and the nucleic acid). There may also be other photophysical properties that are required for the intended purpose of the experiment. For instance, it may be necessary for the fluorophore to be able to photoswitch (a concept that is discussed in detail in Chapter 6).

The fluorophore may be used to detect changes in the surrounding environment rather than changes in the nucleic acid itself. In this case it is advantageous for the reporter to be close to, but not buried within, the duplex. On the other hand, it may be important that the reporter is insensitive to changes in environmental factors such as pH or temperature. This may be necessary for quantitative studies where it is vital that the emission intensity is dependant only on fluorophore concentration. An external

modifier could be designed to be environmentally insensitive but it is also possible that an internal modifier might be intrinsically less sensitive in this regard because of shielding by the surrounding nucleic acid structure. The photostability of a fluorophore may also be improved by relocating it from an external position to an internal position.[110]

Internal modifiers can, potentially, be placed closer to the region (interaction) of interest in the system. There is also more control over the exact positioning (including orientation) of the fluorophore; external modifiers may be attached by flexible linkers that could allow significant deviation in position. As outlined above (§2.1.3.3), knowledge of fluorophore orientation is important for Förster resonance energy transfer (FRET) studies because transfer efficiency is sensitive to dipole alignment. It is also beneficial for modifiers used in FRET studies to be insensitive to environmental factors (including the identity of nearby natural bases) because this would otherwise complicate the interpretation of structural information. It may seem unlikely that internal modifiers could meet all of the requirements to be able to successfully perform FRET studies in nucleic acids but the family of tricyclic cytosine analogues have been shown to be an excellent tool for this purpose;[5,111–113] either of the fluorescent analogues 1,3-diaza-2-oxophenothiazine (tC) or 1,3-diaza-2-oxophenoxazine (tC^O) can be used as a donor with 7-nitro-1,3-diaza-2-oxophenothiazine (tC_{nitro}) acting as an acceptor. The structures of these modifiers are illustrated in Figure 2.12.

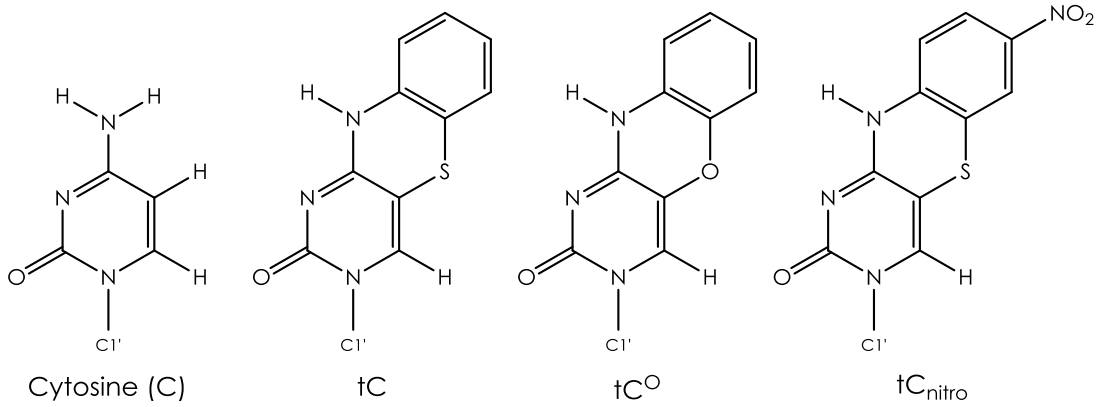

Figure 2.12 – Structures of natural base cytosine (C) and the family of tricyclic cytosine analogues 1,3-diaza-2-oxophenothiazine (tC), 1,3-diaza-2-oxophenoxazine (tC^O) and 7-nitro-1,3-diaza-2-oxophenothiazine (tC_{nitro}). The C1' atom shows the position of the sugar group when the bases are incorporated into a nucleic acid.

External probes may have better photophysical properties in isolation but their superiority may be severely diminished in a nucleic acid; there may be unforeseen

problems such as charge transfer effects or the emergence of mechanisms that lead to DNA damage (for example, radical formation). Internal modifiers are designed for the environment that they are placed in and so should function predictably. The tricyclic cytosine analogues shown above are an example of modifiers that have similar quantum yield when in isolation as a monomer or when incorporated into single-stranded or double-stranded DNA.

External modifiers may be bulky and disrupt normal function of the nucleic acid; of course, despite being better integrated, internal modifiers may also inhibit functionality. The perturbation caused by the insertion of base analogues can be determined in a number of different ways including atomic force microscopy (AFM), nuclear magnetic resonance (NMR), and circular dichroism (CD). Perhaps the most straightforward method is duplex melting temperature studies. Raising the temperature of a duplex in solution eventually causes the two strands to separate (melt) and form single-stranded DNA. The temperature that this melting transition occurs is characteristic of the DNA sequence that is present (higher GC content generally means a higher melting temperature because of more favourable base interactions compared to AT). Comparing the melting temperature of the native duplex with the modified one can therefore give some indication about the structural effect of the modifier. Another method to investigate DNA structure is to perform enzymatic experiments. Restriction enzymes cut particular sequences of DNA and are sensitive to changes in structure. If an enzyme does not perform its usual function with typical efficiency then it can be inferred that the modified nucleic acid has altered physiological properties.

Another consideration when using fluorescent base analogues is the method that is used for incorporation into the nucleic acid. Studies that are focussed on localised changes may only require a single fluorophore in a fairly short oligonucleotide. In such cases it is possible to synthesise a substrate to meet the requirements of the experiment to be performed. This is the approach taken for the dinucleotides that are investigated in Chapter 3. On the other hand, current solid state synthesis methods are limited to producing only short oligonucleotides (100-200 bases) with high fidelity and longer sequences may be required for improved hybridisation (duplex formation) efficiency or structural context, for example. A fully synthetic approach may also become prohibitively expensive for long DNA substrates, especially when tailored fluorescent base analogues are required at high label density. Fortunately, polymerase chain reaction (PCR) can be used to replicate long DNA sequences and, importantly,

certain DNA polymerases are able to tolerate the incorporation of base analogue dyes to create densely labelled DNA sequences.[114–118] This is the approach taken for the study presented in Chapter 6 because it was necessary to have a fluorescent DNA with a high label density. The length of the DNA was not of vital importance for the work carried out in this thesis but future application of the fluorescent DNA produced would likely require a long substrate.

It is worth mentioning that Malyshev *et al.* managed to produce a bacterium that could maintain the presence of an unnatural base pair (UBP) during replication.[119] While the DNA sequence only contained a single UBP, this study offers a tantalising glimpse at what might be possible in the future. If just one of the unnatural bases was fluorescent (unfortunately the photophysics of the unnatural bases was not discussed in the aforementioned paper) then the use of UBPs could enable the investigating the DNA (and associated functions) of living organisms with fluorescence techniques. This being said, base analogues are typically targeted for repair in a replicating system because they pose as undesirable mutations and so it may be very difficult to produce a UBP that is conserved and also has appropriate photophysical properties. There are clearly many factors that must be considered when choosing an appropriate fluorescent base analogue for the experiment at hand.

2.3.3.1 Physiology and Photophysics of 2-Aminopurine

In this thesis, steady-state and time-resolved fluorescence spectroscopy has been enabled by the introduction of internal modifier 2-aminopurine (2AP). 2AP is structurally isomeric to adenine and forms a Watson-Crick base pair with thymine in a duplex (see Figure 2.7 and Figure 2.11). It is also considered an analogue of guanine (2AP lacks only H1 and O6 atoms) and can form a wobble base pair with cytosine. In fact, during structural analysis, 2AP is sometimes considered a substitute of guanine, rather than adenine, due to their more similar geometry. The ability of 2AP to mimic natural base interactions means that it can be assimilated into a duplex with minimal perturbation to the structure.

Despite being structurally very similar to the natural purine bases 2AP has profoundly different fluorescence character. The quantum yield (Φ, ~0.6) and fluorescence lifetime (τ, 11.5 ns) of 2AP are considerably greater than that of the natural bases (Φ, $\sim 10^{-4}$; τ, sub-ps).[120,121] An early study of 2-substituted purine structures concluded that the distinct fluorescence character of 2AP compared to the natural bases was due to a change in electronic structure; specifically there was a

reversal of the energy of the nπ* and ππ* excited states.[122] The electron-donating NH_2 group at the C2 position is thought to disrupt the linear combination of non-bonding orbitals of the endocyclic N1 and N3 atoms and increase the energy of the nπ* transition. It is worth noting that the presence of a substituent on the C6 atom is enough to (at least partly) counteract the electronic inversion; this is especially true for an oxygen atom because it can promote intersystem crossing.[122] This could explain why guanine, which also has an NH_2 group at the C2 position, is essentially non-fluorescent like the other natural bases; however, a more significant factor is likely to be the presence of the hydrogen atom at N1, which disrupts the conjugated bonding of the 6-membered ring.

At this point it is worth highlighting that solvation plays a pivotal role in determining the fluorescence character of 2AP. In the gas-phase 2AP is, like the natural bases, essentially non-fluorescent.[123,124] Interestingly, recent microhydration experiments have shown that even a single water molecule may cause a dramatic change to the fluorescence properties of 2AP;[125] however, despite much effort, the precise nature of the photophysical properties of 2AP is still not fully understood. While the non-fluorescent nature of gas-phase 2AP is worth bearing in mind, it should be assumed that discussion within this thesis relates to the properties of solvated 2AP.

The absorption maximum of 2AP (~305 nm) is significantly red-shifted compared to the natural bases (~260 nm) which means that it can be selectively excited during experiments. This is especially useful when 2AP is in a system where it is vastly outnumbered by natural bases (such as a duplex).[126] Additionally, as highlighted below, 2AP fluorescence is extremely sensitive to its local environment and incorporation into a nucleic acid construct causes around a 100-fold decrease in quantum yield. The distinct fluorescence properties of 2AP have made it a valuable asset in the study of DNA conformation and dynamics. Table 2.2 shows a comparison of typical time-resolved parameters (lifetimes, τ_i, and their associated probability amplitudes, A_i) obtained for 2AP when in a variety of nucleic acid systems.

2AP System	τ_1 /ns	τ_2 /ns	τ_3 /ns	τ_4 /ns	A_1	A_2	A_3	A_4
Solution (H_2O)	-	-	-	11.5	-	-	-	1.00
Dinucleotide	< 0.1	0.2-0.6	2.0-4.0	7.0-9.0	0.40	0.15	0.40	0.05
ssDNA					0.50	0.20	0.15	0.15
dsDNA					0.80	0.10	0.05	0.05

Table 2.2 – Comparison of typical fluorescence lifetimes (τ_i) and A-factors (A_i) obtained for 2AP in solution, dinucleotides, single-stranded DNA (ssDNA), and double-stranded DNA (dsDNA) using TCSPC. A-factors represent the fractional contribution to the overall fluorescence decay curve. The values shown are based on the work of Somsen *et al.,*[127] Neely *et al.,*[128–130] and Wu.[131]

2AP in solution is found to have a single fluorescence decay lifetime of around 11.5 ns (neglecting the effect of tautomerism).[128] In contrast, when 2AP is present in a nucleic acid construct it displays a far more complicated behaviour and typically four lifetimes are required to adequately fit the fluorescence decay data. Furthermore, these lifetime values span two orders of magnitude indicating that 2AP experiences a very heterogeneous environment. A common interpretation of these observations is that 2AP is occupying a number of distinct conformational states, each of which has a different set of intermolecular interactions associated with it. The shortest lifetime (which is usually found to be less than 100 ps) is thought to be due to 2AP which is integrated into the highly organised structure of the DNA double helix. Charge transfer with proximal bases has been implicated as an effective method by which excited state relaxation can occur on the sub-ps timescale.[24,132] The longest lifetime component (~9 ns) has been attributed to extra-helical 2AP. In this conformation 2AP is essentially isolated from the rest of the structure and only interacts with solvent. Access to this state can be achieved by rotation out of the double helix in a process known as base-flipping.[133] The transition between stacked and flipped states is discussed in more detail later. The slight reduction in the lifetime value of extra-helical 2AP compared to when 2AP is truly isolated in solution can partly be put down to the fact that 2AP is covalently bonded to a (deoxy)ribose sugar group. The fluorescence decay of 2AP riboside in solution can be modelled with a single lifetime component of about 10.5 ns.[128] Collisional quenching with the surrounding structure might also be responsible for reducing the lifetime. It has been more of a challenge to identify conformational states that could be responsible for the two intermediate lifetimes (~0.3 ns and ~3 ns). It is perhaps possible that they are due to geometries in which 2AP is only partially stacked but still (mainly) in the hydrophobic environment of the

double helix. There have also been proposals of dark states[120,134] and gated relaxation[127] as an explanation of these lifetimes.

Despite the complexity of the fluorescence decay of 2AP, lifetime values are essentially unchanged between dinucleotides, single-stranded DNA (ssDNA), and double-stranded DNA (dsDNA). This indicates that 2AP can sample similar environments in all of these systems. The values in Table 2.2 show that while lifetimes remain almost constant there are changes in the relative occupancy of the associated states. For example, the proportion of 2AP which undergoes rapid relaxation (A_1) increases as the system becomes more constrained to the canonical duplex structure (dinucleotide → ssDNA → dsDNA). This increase can be attributed to greater charge transfer efficiency as stacking between bases becomes more stable. In a dinucleotide, 2AP has only one neighbouring base to interact with. In contrast, when 2AP is in a DNA construct it is sandwiched between two bases and, in the case of a duplex, also has further restriction imposed by base-pairing.

Given the difference in the structural constraints between the various systems it is actually somewhat surprising that the changes in relative populations are not significantly greater. It is particularly interesting that the occupancy of the longest lifetime component – where 2AP is thought to be unstacked (base-flipped) – shows little change between dinucleotide and DNA. This suggests that, even in a dinucleotide, there is a considerable preference for the 2AP to be interacting in some way with the rest of the structure rather than being isolated in solvent.

Table 2.2 is, of course, just a simplistic summary of *typical* parameters. Specific samples can have different behaviour depending on their particular construct. For example, DNA sequence plays an important role in determining the precise photophysical behaviour of the incorporated 2AP fluorophore. This is particularly true in the case of dinucleotides where there is only a single base neighbouring 2AP. It is also important to consider the time-resolution of the experiments used to obtain the above values; decay processes which occur faster than ~50 ps are not detectable using time-correlated single photon counting (TCSPC) and so it is possible that some differences are obscured by the experimental method. The causes of the limit to the temporal resolution of a TCSPC experiment are discussed in Chapter 4.

Regardless of these caveats, the values show that there are some interesting similarities and differences between the photophysical properties of 2AP when incorporated into dinucleotide or double-helical constructs. Understanding the cause of this behaviour is essential if an accurate description of these systems is to be obtained;

Chapter 3 aims to address this issue. This knowledge should be of great interest to a wide audience given the extensive use of 2AP in studying a vast array of biophysical systems.

One particularly pertinent example of the use of 2AP in the study of DNA is in the investigation of base-flipping. The duplex shown in Figure 2.4 (§2.3.1) was taken from a complex containing methyltransferase *Haemophilus haemolyticus* I (M. HhaI). For methylation to occur the target base (in this case cytosine) is required to enter a catalytic pocket within the enzyme. Nature has developed an elegant method to achieve this feat whereby rotation out of the double-helical structure (base-flipping) allows capture of the target base by the enzyme. In some cases the remaining duplex structure is stabilised by closure of an enzymatic loop into the nascent space left behind by the flipped base. It should be noted that, for clarity of the DNA structure, the M. HhaI enzyme has been omitted from Figure 2.4. Base-flipping therefore causes a dramatic conformational change to duplex structure by disrupting base-stacking continuity. Clearly the local environment of the target base and also proximal bases changes significantly during this process and so base-flipping is highly suited to being studied by 2AP fluorescence. 2AP can either be incorporated as the target base itself or placed in a proximal location which is also perturbed.

It is worth pointing out that bases can spontaneously flip out of a duplex in a process known as DNA breathing.[135] 2AP is therefore able to access a state which is essentially interaction-free even without the presence of a base-flipping enzyme. DNA breathing is very likely to be the cause of the small population of DNA-incorporated 2AP which exhibits a solution-like fluorescence lifetime.

2.4 References

1. J. M. Hollas, *Modern Spectroscopy*, John Wiley & Sons, Ltd, 4th edn., 2004.

2. P. Atkins and J. de Paula, *Atkins' Physical Chemistry*, Oxford University Press, 8th edn., 2006.

3. N. J. Turro, *Modern Molecular Photochemistry*, University Science Books, 1991.

4. J. R. Lackowicz, *Principles of Fluorescence Spectroscopy*, Springer Science & Business Media, LLC, 3rd edn., 2006.

5. K. Börjesson, S. Preus, A. H. El-Sagheer, T. Brown, B. Albinsson, and L. M. Wilhelmsson, *J. Am. Chem. Soc.*, 2009, **131**, 4288–93.

6. L. M. Wilhelmsson, *Q. Rev. Biophys.*, 2010, **43**, 159–83.

7. J. Ouellet, S. Schorr, A. Iqbal, T. J. Wilson, and D. M. J. Lilley, *Biophys. J.*, 2011, **101**, 1148–54.

8. A. Iqbal, S. Arslan, B. Okumus, T. J. Wilson, G. Giraud, D. G. Norman, T. Ha, and D. M. J. Lilley, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 11176–81.

9. D. V. O'Connor and D. Phillips, *Time-Correlated Single Photon Counting*, Academic Press, 1984.

10. W. Becker, *Advanced Time-Correlated Single Photon Counting Techniques*, Springer Berlin Heidelberg, 2005.

11. F. Jensen, *Introduction to Computational Chemistry*, John Wiley & Sons, Ltd, 2nd edn., 2007.

12. D. C. Young, *Computational Chemistry - A Practical Guide for Applying Techniques to Real World Problems*, John Wiley & Sons, Inc., 2001.

13. D. Gibson, PhD Thesis, University of Edinburgh, 2007.

14. P. Hohenberg and W. Kohn, *Phys. Rev.*, 1964, **136**, B864–71.

15. W. Kohn and L. J. Sham, *Phys. Rev.*, 1965, **140**, A1133–8.

16. Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215–41.

17. Y. Zhao and D. G. Truhlar, *Acc. Chem. Res.*, 2008, **41**, 157–67.

18. Y. Zhao and D. G. Truhlar, *Chem. Phys. Lett.*, 2011, **502**, 1–13.

19. R. Ditchfield, W. J. Hehre, and J. A. Pople, *J. Chem. Phys.*, 1971, **54**, 724–8.

20. R. S. Hunter and T. van Mourik, *J. Comput. Chem.*, 2012, **33**, 2161–72.

21. S. Miertuš, E. Scrocco, and J. Tomasi, *Chem. Phys.*, 1981, **55**, 117–29.

22. IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN), *Eur. J. Biochem.*, 1983, **131**, 9–15.

23. M. O'Gara, S. Klimašauskas, R. J. Roberts, and X. Cheng, *J. Mol. Biol.*, 1996, **261**, 634–45.

24. S. O. Kelley and J. K. Barton, *Science*, 1999, **283**, 375–81.

25. M. A. O'Neill, H.-C. Becker, C. Wan, J. K. Barton, and A. H. Zewail, *Angew. Chem. Int. Ed. Engl.*, 2003, **42**, 5896–900.

26. International Human Genome Sequencing Consortium, *Nature*, 2004, **50**, 162–8.

27. S. G. Gregory, K. F. Barlow, K. E. McLay, R. Kaul, D. Swarbreck, A. Dunham, C. E. Scott, K. L. Howe, K. Woodfine, C. C. A. Spencer, M. C. Jones, C. Gillson, S. Searle, Y. Zhou, F. Kokocinski, L. McDonald, R. Evans, K. Phillips, A. Atkinson, R. Cooper, C. Jones, R. E. Hall, T. D. Andrews, C. Lloyd, R. Ainscough, J. P. Almeida, K. D. Ambrose, F. Anderson, R. W. Andrew, R. I. S. Ashwell, K. Aubin, A. K. Babbage, C. L. Bagguley, J. Bailey, R. Banerjee, H. Beasley, G. Bethel, C. P. Bird, S. Bray-Allen, J. Y. Brown, A. J. Brown, S. P. Bryant, D. Buckley, D. C. Burford, W. D. H. Burrill, J. Burton, J. Bye, C. Carder, J. C. Chapman, S. Y. Clark, G. Clarke, C. Clee, S. M. Clegg, V. Cobley, R. E. Collier, N. Corby, G. J. Coville, J. Davies, R. Deadman, P. Dhami, O. Dovey, M. Dunn, M. Earthrowl, A. G. Ellington, H. Errington, L. M. Faulkner, A. Frankish, J. Frankland, L. French, P. Garner, J. Garnett, L. Gay, M. R. J. Ghori, R. Gibson, L. M. Gilby, W. Gillett, R. J. Glithero, D. V. Grafham, S. M. Gribble, C. Griffiths, S. Griffiths-Jones, R. Grocock, S. Hammond, E. S. I. Harrison, E. Hart, E. Haugen, P. D. Heath, S. Holmes, K. Holt, P. J. Howden, A. R. Hunt, S. E. Hunt, G. Hunter, J. Isherwood, R. James, C. Johnson, D. Johnson, A. Joy, M. Kay, J. K. Kershaw, M. Kibukawa, A. M. Kimberley, A. King, A. J. Knights, H. Lad, G. Laird, C. F. Langford, S. Lawlor, D. A. Leongamornlert, D. M. Lloyd, J. Loveland, J. Lovell, M. J. Lush, R. Lyne, S. Martin, M. Mashreghi-Mohammadi, L. Matthews, N. S. W. Matthews, S. McLaren, S. Milne, S. Mistry, M. J. F. Moore, T. Nickerson, C. N. O'Dell, K. Oliver, A. Palmeiri, S. A. Palmer, R. D. Pandian, A. Parker, D. Patel, A. V. Pearce, A. I. Peck, S. Pelan, K. Phelps, B. J. Phillimore, R. Plumb, K. M. Porter, E. Prigmore, J. Rajan, C. Raymond, G. Rouse, C. Saenphimmachak, H. K. Sehra, E. Sheridan, R. Shownkeen, S. Sims, C. D. Skuce, M. Smith, C. Steward, S. Subramanian, N. Sycamore, A. Tracey, A. Tromans, Z. Van Helmond, M. Wall, J. M. Wallis, S. White, S. L. Whitehead, J. E. Wilkinson, D. L. Willey, H. Williams, L. Wilming, P. W. Wray, Z. Wu, A. Coulson, M. Vaudin, J. E. Sulston, R. Durbin, T. Hubbard, R. Wooster, I. Dunham, N. P. Carter, G. McVean, M. T. Ross, J. Harrow, M. V. Olson, S. Beck, J. Rogers, and D. R. Bentley, *Nature*, 2006, **441**, 315–21.

28. H. M. Berman and W. K. Olson, in *DNA 50: The Secret of Life*, eds. P. Antell and M. Balaban, Faircout Limited, London, 2003, pp. 104–24.

29. R. Dahm, *Hum. Genet.*, 2008, **122**, 565–81.

30. F. Griffith, *J. Hyg. (Lond).*, 1928, **XXVII**, 113–59.

31. O. T. Avery, C. M. MacLeod, and M. McCarty, *J. Exp. Med.*, 1944, **79**, 137–59.

32. A. D. Hershey and M. Chase, *J. Gen. Physiol.*, 1952, **36**, 39–56.

33. M. E. Jones, *Yale J. Biol. Med.*, 1953, **26**, 80–97.

34. P. A. Levene, *J. Biol. Chem.*, 1919, **40**, 415–24.

35. W. T. Astbury, *Symp. Soc. Exp. Biol.*, 1947, **1**, 66–76.

36. L. Pauling and R. B. Corey, *Proc. Natl. Acad. Sci. U. S. A.*, 1953, **39**, 84–97.

37. S. Furberg, *Acta Chem. Scand.*, 1952, **6**, 634–40.

38. J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 737–8.

39. M. H. F. Wilkins, A. R. Stokes, and H. R. Wilson, *Nature*, 1953, **171**, 738–40.

40. R. E. Franklin and R. G. Gosling, *Nature*, 1953, **171**, 740–1.

41. R. E. Franklin and R. G. Gosling, *Nature*, 1953, **172**, 156–7.

42. J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 964–7.

43. E. Chargaff, E. Vischer, R. Doniger, C. Green, and F. Misani, *J. Biol. Chem.*, 1949, **177**, 405–16.

44. S. Zamenhof, G. Brawerman, and E. Chargaff, *Biochim. Biophys. Acta*, 1952, **9**, 402–5.

45. G. R. Wyatt, *J. Gen. Physiol.*, 1952, **36**, 201–5.

46. V. N. Soyfer, *Nat. Rev. Genet.*, 2001, **2**, 723–9.

47. R. E. Franklin and R. G. Gosling, *Acta Crystallogr.*, 1953, **6**, 673–7.

48. A. G. W. Leslie, S. Arnott, R. Chandrasekaran, and R. L. Ratliff, *J. Mol. Biol.*, 1980, **143**, 49–72.

49. A. Ghosh and M. Bansal, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2003, **59**, 620–6.

50. M. Bansal, *Curr. Sci.*, 2003, **85**, 1556–63.

51. K. Hoogsteen, *Acta Crystallogr.*, 1959, **12**, 822–3.

52. K. Hoogsteen, *Acta Crystallogr.*, 1963, **16**, 907–16.

53. J. Rubin, T. Brennan, and M. Sundaralingam, *Science*, 1971, **174**, 1020–2.

54. N. C. Seeman, J. L. Sussman, H. M. Berman, and S. H. Kim, *Nat. New Biol.*, 1971, **233**, 90–2.

55. J. Rubin, T. Brennan, and M. Sundaralingam, *Biochemistry*, 1972, **11**, 3112–28.

56. J. L. Sussman, N. C. Seeman, S.-H. Kim, and H. M. Berman, *J. Mol. Biol.*, 1972, **66**, 403–21.

57. N. Camerman, J. K. Fawcett, and A. Camerman, *Science*, 1973, **182**, 1142–3.

58. R. O. Day, N. C. Seeman, J. M. Rosenberg, and A. Rich, *Proc. Natl. Acad. Sci. U. S. A.*, 1973, **70**, 849–53.

59. J. M. Rosenberg, N. C. Seeman, J. J. P. Kim, F. L. Suddath, H. B. Nicholas, and A. Rich, *Nature*, 1973, **243**, 150–4.

60. N. C. Seeman, J. M. Rosenberg, F. L. Sluddath, J. J. P. Kim, and A. Rich, *J. Mol. Biol.*, 1976, **104**, 109–44.

61. J. M. Rosenberg, N. C. Seeman, R. O. Day, and A. Rich, *J. Mol. Biol.*, 1976, **104**, 145–67.

62. W. B. T. Cruse, E. Egert, O. Kennard, G. B. Sala, S. A. Salisbury, and M. A. Viswamitra, *Biochemistry*, 1983, **22**, 1833–9.

63. M. Coll, X. Solans, M. Font-Altaba, and J. A. Subirana, *J. Biomol. Struct. Dyn.*, 1987, **4**, 797–811.

64. Y. Inoue and S. Aoyagi, *Biochem. Biophys. Res. Commun.*, 1967, **28**, 973–7.

65. S. I. Chan and J. H. Nelson, *J. Am. Chem. Soc.*, 1969, **65**, 168–83.

66. B. W. Bangerter and S. I. Chan, *J. Am. Chem. Soc.*, 1969, **91**, 3910–21.

67. P. O. P. Ts'o, N. S. Kondo, M. P. Schweizer, and P. Hollis, Donald, *Biochemistry*, 1969, **8**, 997–1029.

68. D. J. Wood, K. K. Ogilvie, and F. E. Hruska, *Can. J. Chem.*, 1975, **53**, 2781–90.

69. S.-H. Kim, H. M. Berman, N. C. Seeman, and M. D. Newton, *Acta Crystallogr. Sect. B Struct. Crystallogr. Cryst. Chem.*, 1973, **29**, 703–10.

70. B. Schneider, Z. Morávek, and H. M. Berman, *Nucleic Acids Res.*, 2004, **32**, 1666–77.

71. A. Barth and P. I. Haris, in *Biological and Biomedical Infrared Spectroscopy*, IOS Press BV, 2009, pp. 1–53.

72. E. R. Blout and M. Fields, *Science*, 1948, **107**, 252.

73. M. J. Fraser and R. D. B. Fraser, *Nature*, 1951, **167**, 761–2.

74. E. R. Blout and H. Lenormant, *J. Opt. Soc. Am.*, 1953, **43**, 1093–5.

75. M. Falk, K. A. Hartman, Jr., and R. C. Lord, *J. Am. Chem. Soc.*, 1963, **39**, 9–13.

76. G. J. Thomas, Jr., *Biopolymers*, 1969, **7**, 325–34.

77. T. Elsaesser, in *Ultrafast Infrared Vibrational Spectroscopy*, ed. M. D. Fayer, CRC Press, 2013, pp. 35–72.

78. G. W. Doorley, D. A. McGovern, M. W. George, M. Towrie, A. W. Parker, J. M. Kelly, and S. J. Quinn, *Angew. Chem. Int. Ed. Engl.*, 2009, **48**, 123–7.

79. M. Towrie, G. W. Doorley, M. W. George, A. W. Parker, S. J. Quinn, and J. M. Kelly, *Analyst*, 2009, **134**, 1265–73.

80. P. M. Keane, M. Wojdyla, G. W. Doorley, J. M. Kelly, I. P. Clark, A. W. Parker, G. M. Greetham, M. Towrie, L. M. Magno, and S. J. Quinn, *Phys. Chem. Chem. Phys.*, 2012, **14**, 6307–11.

81. G. W. Doorley, M. Wojdyla, G. W. Watson, M. Towrie, A. W. Parker, J. M. Kelly, and S. J. Quinn, *J. Phys. Chem. Lett.*, 2013, **4**, 2739–44.

82. R. Ahmad, H. Arakawa, and H. A. Tajmir-Riahi, *Biophys. J.*, 2003, **84**, 2460–6.

83. G. A. Rodley, R. S. Scobie, R. H. Bates, and R. M. Lewitt, *Proc. Natl. Acad. Sci. U. S. A.*, 1976, **73**, 2959–63.

84. V. Amarnath and A. D. Broom, *Chem. Rev.*, 1977, **77**, 183–217.

85. M. Caruthers, *Acc. Chem. Res.*, 1991, **24**, 278–84.

86. A. H.-J. Wang, G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. van Boom, G. van der Marel, and A. Rich, *Nature*, 1979, **282**, 680–6.

87. R. E. Dickerson, H. R. Drew, B. N. Conner, R. M. Wing, A. V. Fratini, and M. L. Kopka, *Science*, 1982, **216**, 475–85.

88. R. Wing, H. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura, and R. E. Dickerson, *Nature*, 1980, **287**, 755–8.

89. H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, and R. E. Dickerson, *Proc. Natl. Acad. Sci. U. S. A.*, 1981, **78**, 2179–83.

90. Z. Shakked, D. Rabinovich, W. B. T. Cruse, E. Egert, O. Kennard, G. Sala, S. A. Salisbury, and M. A. Viswamitra, *Proc. R. Soc. B Biol. Sci.*, 1981, **213**, 479–87.

91. J. C. Wang, *Proc. Natl. Acad. Sci. U. S. A.*, 1979, **76**, 200–3.

92. W. Kabsch, C. Sander, and E. N. Trifonov, *Nucleic Acids Res.*, 1982, **10**, 1097–104.

93. B. H. M. Mooers, *Methods*, 2009, **47**, 168–76.

94. H. M. Berman, W. K. Olson, D. L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S.-H. Hsieh, A. R. Srinivasan, and B. Schneider, *Biophys. J.*, 1992, **63**, 751–9.

95. B. Coimbatore Narayanan, J. Westbrook, S. Ghosh, A. I. Petrov, B. Sweeney, C. L. Zirbel, N. B. Leontis, and H. M. Berman, *Nucleic Acids Res.*, 2014, **42**, D114–22.

96. A. Dallmann, L. Dehmel, T. Peters, C. Mügge, C. Griesinger, J. Tuma, and N. P. Ernsting, *Angew. Chem. Int. Ed. Engl.*, 2010, **49**, 5989–92.

97. Protein Data Bank, *Nat. New Biol.*, 1971, **233**, 223.

98. F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, *J. Mol. Biol.*, 1977, **112**, 535–42.

99. H. M. Berman, G. J. Kleywegt, H. Nakamura, and J. L. Markley, *Structure*, 2012, **20**, 391–6.

100. W. C. Johnson, Jr. and I. Tinoco, Jr., *Biopolymers*, 1969, **7**, 727–49.

101. C. R. Cantor, M. M. Warshaw, and H. Shapiro, *Biopolymers*, 1970, **9**, 1059–77.

102. N. P. Johnson and E. Switkes, *Biopolymers*, 1978, **17**, 857–72.

103. S. M. Law, R. Eritja, M. F. Goodman, and K. J. Breslauer, *Biochemistry*, 1996, **35**, 12329–37.

104. N. P. Johnson, W. A. Baase, and P. H. von Hippel, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 3426–31.

105. J. P. Peters, L. S. Mogil, M. J. McCauley, M. C. Williams, and L. J. Maher, III, *Biophys. J.*, 2014, **107**, 448–59.

106. J. Gidden and M. T. Bowers, *Eur. Phys. J. D*, 2002, **20**, 409–19.

107. D. P. Millar, *Curr. Opin. Struct. Biol.*, 1996, **6**, 322–6.

108. J. N. Wilson and E. T. Kool, *Org. Biomol. Chem.*, 2006, **4**, 4265–74.

109. R. W. Sinkeldam, N. J. Greco, and Y. Tor, *Chem. Rev.*, 2010, **110**, 2579–619.

110. W. Lee, P. H. von Hippel, and A. H. Marcus, *Nucleic Acids Res.*, 2014, **42**, 5967–77.

111. L. M. Wilhelmsson, A. Holme, P. Lincoln, P. E. Nielsen, and B. Norde, *J. Am. Chem. Soc.*, 2001, **123**, 2434–5.

112. P. Sandin, K. Börjesson, H. Li, J. Mårtensson, T. Brown, L. M. Wilhelmsson, and B. Albinsson, *Nucleic Acids Res.*, 2008, **36**, 157–67.

113. S. Preus, K. Bo, K. Kilså, B. Albinsson, and L. M. Wilhelmsson, *J. Phys. Chem. B*, 2010, **114**, 1050–6.

114. G. Giller, T. Tasara, B. Angerer, K. Mühlegger, M. Amacker, and H. Winter, *Nucleic Acids Res.*, 2003, **31**, 2630–5.

115. T. Tasara, B. Angerer, M. Damond, H. Winter, S. Dörhöfer, U. Hübscher, and M. Amacker, *Nucleic Acids Res.*, 2003, **31**, 2636–46.

116. J. P. Anderson, B. Angerer, and L. A. Loeb, *Biotechniques*, 2005, **38**, 257–64.

117. G. Stengel, M. Urban, B. W. Purse, and R. D. Kuchta, *Anal. Chem.*, 2009, **81**, 9079–85.

118. N. Ramsay, A.-S. Jemth, A. Brown, N. Crampton, P. Dear, and P. Holliger, *J. Am. Chem. Soc.*, 2010, **132**, 5096–104.

119. D. A. Malyshev, K. Dhami, T. Lavergne, T. Chen, N. Dai, J. M. Foster, I. R. Corrêa, Jr., and F. E. Romesberg, *Nature*, 2014, **509**, 385–8.

120. O. F. A. Larsen, I. H. M. van Stokkum, F. L. de Weerd, M. Vengris, C. T. Aravindakumar, R. van Grondelle, N. E. Geacintov, and H. van Amerongen, *Phys. Chem. Chem. Phys.*, 2004, **6**, 154–60.

121. O. J. G. Somsen, A. van Hoek, and H. van Amerongen, *Chem. Phys. Lett.*, 2005, **402**, 61–5.

122. J. Drobnik and L. Augenstein, *Photochem. Photobiol.*, 1966, **5**, 83–97.

123. K. A. Seefeld, C. Plützer, D. Löwenich, T. Häber, R. Linder, K. Kleinermanns, J. Tatchen, and C. M. Marian, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3021–6.

124. K. Feng, G. Engler, K. Seefeld, and K. Kleinermanns, *Chemphyschem*, 2009, **10**, 886–9.

125. S. Lobsiger, S. Blaser, R. K. Sinha, H.-M. Frey, and S. Leutwyler, *Nat. Chem.*, 2014, **6**, 989–93.

126. D. C. Ward, E. Reich, and L. Stryer, *J. Biol. Chem.*, 1969, **244**, 1228–37.

127. O. J. G. Somsen, L. B. Keukens, M. N. de Keijzer, A. van Hoek, and H. van Amerongen, *Chemphyschem*, 2005, **6**, 1622–7.

128. R. K. Neely, S. W. Magennis, D. T. F. Dryden, and A. C. Jones, *J. Phys. Chem. B*, 2004, **108**, 17606–10.

129. R. K. Neely, D. Daujotyte, S. Grazulis, S. W. Magennis, D. T. F. Dryden, S. Klimasauskas, and A. C. Jones, *Nucleic Acids Res.*, 2005, **33**, 6953–60.

130. R. K. Neely, G. Tamulaitis, K. Chen, M. Kubala, V. Siksnys, and A. C. Jones, *Nucleic Acids Res.*, 2009, **37**, 6859–70.

131. X. Wu, PhD Thesis, University of Edinburgh, 2012.

132. T. Fiebig, C. Wan, and A. H. Zewail, *Chemphyschem*, 2002, **3**, 781–8.

133. X. Cheng and R. M. Blumenthal, *Structure*, 1996, **4**, 639–45.

134. O. F. A. Larsen, I. H. M. van Stokkum, M.-L. Groot, J. T. M. Kennis, R. van Grondelle, and H. van Amerongen, *Chem. Phys. Lett.*, 2003, **371**, 157–63.

135. P. H. von Hippel, N. P. Johnson, and A. H. Marcus, *Biopolymers*, 2013, **99**, 923–54.

Chapter 3: Time-Resolved Fluorescence of 2AP-Containing Dinucleotides in Alcohol-Water Mixtures

Evans boldly put 50 atm of ethylene [...] in a cell with 25 atm of O_2. The apparatus subsequently blew up, but luckily not before he had obtained the spectra shown in Figure 8.

– Merer and Mulliken, *Chem. Rev.,* 1969, **69,** 645

3.1 Introduction

As outlined in Chapter 2, fluorescent base-analogue 2-aminopurine (2AP) has become a valuable asset in the study of nucleic acids due to its highly desirable structural and photophysical properties; however, there still remains a significant shortfall in the understanding of the precise causes of its complex fluorescence behaviour within these systems. Of particular concern is the inability to provide a definitive physical model that can fully explain the multi-exponential fluorescence decay observed from time-resolved fluorescence measurements of 2AP incorporated in oligonucleotides. This behaviour contrasts greatly from the mono-exponential decay of free 2AP in solution and typically four-exponentials are required to adequately fit the decays measured within the picosecond-nanosecond time-range. These four components are generally attributed to distinct conformational states in which 2AP experiences varying degrees of intermolecular interactions. Charge transfer between efficiently stacked bases has been implicated as the cause of the shortest lifetime (τ_1, < 100 ps) observed. The longest lifetime (τ_4, ~9 ns) is similar to that of free 2AP and is commonly accredited to a conformation where the 2AP moiety does not experience significant intramolecular interaction. Such a state is thought to be accessible in DNA through base-flipping; as highlighted in the previous chapter. The two intermediate lifetimes (~0.5 ns and ~2 ns) have been more controversial, however, and there is still a lack of knowledge regarding their precise origin. It is possible that they are due to intermediate conformations between fully stacked and open forms but alternative explanations, such as dark states, have also been proposed.

In this chapter, 2AP-containing dinucleotides are investigated using steady-state and time-resolved fluorescence spectroscopy in an attempt to clarify factors which govern relaxation pathway dynamics of 2AP in nucleic acids. Dinucleotides offer a simple model of DNA by eliminating all but one of the bases neighbouring 2AP; this provides a means to identify base-specific effects and greatly simplifies data interpretation. Experiments have been performed in alcohol-water mixtures to systematically investigate the effects of changing the solvent environment on the observed photophysical behaviour. To aid in the understanding of the interpretation of the results, a brief summary of recent literature will now be discussed.

Reichardt *et al.* used broadband transient absorption spectroscopy to investigate 2AP-deoxyriboside in various solvents;[1] their conclusions provide the current interpretation of solvation effects on 2AP photophysics. The use of the deoxyriboside

meant that the system was more akin to the moiety present in nucleic acids and also circumvented potential complications caused by tautomerism between the 9H- and 7H-forms of 2AP.[2] A comparison of the structures of the different tautomers of 2AP, 2AP-riboside, and 2AP-deoxyriboside is shown in Figure 3.1.

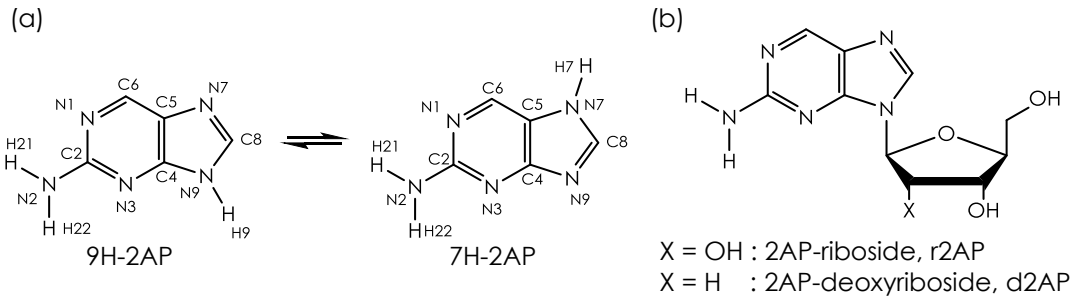

Figure 3.1 – Molecular structure of (a) the 9H- and 7H-forms of 2AP and (b) 2AP-riboside and 2AP-deoxyriboside. In water the 7H-forms of 2AP accounts for around 40% of the total population.[2]

Transient absorption is a pump-probe technique that follows the temporal evolution of the changes in the absorption profile of a sample during relaxation after pulsed excitation. Importantly, this technique can detect states which have either large or small radiative transition cross-sections and so the presence of dark states can be directly measured (and quantified). Broadband detection can be used to provide simultaneous measurement of the whole spectral response allowing the use of global analysis, which is much more robust than single-wavelength analysis. Transient absorption measurements are complicated by the fact that the measured signal has contributions from ground state bleaching (GSB), excited-state absorption (ESA), stimulated emission (SE), and potentially photoproduct absorption.[3] Analysis of the data can therefore become ambiguous and it may be possible for a number of different models to fit the data with similar quality. Transient absorption does not necessarily report on the initially populated, emitting state, if this decays rapidly to a dark state; it is possible that the fate of the emitting state may only be revealed fully through ultrafast fluorescence measurements (upconversion or downconversion) in combination with transient absorption.

The results obtained by Reichardt *et al.* suggest that ~10% to 40% of the initial excited-state population of 2AP decays through a triplet state ($^3\pi\pi^*$). Formation of the triplet state was tentatively proposed to occur *via* internal conversion (IC) to the $^1n\pi^*$ state ($^3\pi\pi^* \leftarrow {}^1n\pi^* \leftarrow {}^1\pi\pi^*$). The barrier height to IC was found to be dependent on the polarity of the solvent as well as its H-bonding ability; however, as the results were

based on the measurements of only three solvents which all varied in both of these properties, it is not possible to extract the relative contribution of these factors. Interestingly, the combined fluorescence and phosphorescence yields accounted for ~70% of the total excited-state population and this was independent of solvent type; although, the authors conceded that further solvents needed to be examined before the generality of this effect was confirmed.

Larsen *et al.* used transient absorption to study free 2AP and 2AP incorporated into dinucleotides and DNA.[4,5] Their results indicated the presence of a dark state that accounted for around 30% of the total excited-state population of free 2AP, which is in agreement with the conclusions of Reichardt *et al*;[1] however, a significant shortfall in the measurements made was the inability to accurately determine the long lifetime states. The time-range used (1 ns or 5 ns) was only a fraction of the longest lifetime typically observed for 2AP and it was not possible to accurately characterise this component. Given the interdependence of fitting parameters it seems unlikely that an accurate representation of the system was achieved. Indeed, their analysis of free 2AP gave lifetime components of 1.3 ps, 3.5 ns, and ~14 ns and while the shortest lifetime could be a result of solvation dynamics (consistent with Reichardt *et al.*[1]) the latter two lifetimes would appear to be an artificial splitting of the typically measured lifetime of ~10 ns. The heterogeneity of the dinucleotides and duplex constructs means that analysis of the associated decays is far more complicated than for free 2AP. Therefore these results will not be discussed here because of the aforementioned ambiguity caused by using a short time-range during experiments.

Charge transfer dynamics in complexes of 2AP and dNTPs (deoxynucleotide triphosphates) has been investigated using transient absorption and fluorescence upconversion by Fiebig, Wan, and Zewail.[6] Their results implicate both electron transfer (ET) and hole transfer (HT) as an efficient quenching mechanism in these systems for excited-state 2AP (2AP*). To be clear, ET relates to oxidation of 2AP* while HT relates to reduction of 2AP*.

Free energy estimates based on measured decay rates implied that HT occurs to guanine, G, and adenine, A, (for example, $2AP^* + G \rightarrow 2AP^{\bullet-} + G^{\bullet+}$), while ET occurs to thymine, T, and cytosine, C, (for example, $2AP^* + T \rightarrow 2AP^{\bullet+} + T^{\bullet-}$); the process efficiency was predicted to be greater for the former base in both cases. It was concluded that either redox process could potentially occur for A and hypoxanthine (inosine), I. The results are in line with the relative propensity for the natural bases to

be oxidised (G > A > C > T) or reduced (T > C > A > G); although, these trends were obtained in acetonitrile (ACN) and dimethylformamide (DMF), respectively.[6,7]

A comparison of the measured lifetime of the fast decay component and estimates of the reduction potential for the different bases is shown in Table 3.1. Although they are not strictly correlated, it is evident that the redox properties of the neighbouring base play a role in determining the observed lifetime. It should be noted that oxidation, rather than reduction potentials, should be used for determining ET efficacy but there is currently a lack of data available for aqueous solutions.

Nucleotide	Reduction Potential /V (versus NHE)	Δ(2AP*) /V	Lifetime /ps
Guanosine (G)	+1.29	−0.21	19
Adenosine (A)	+1.42	−0.08	350
2AP*	+1.50	0.00	-
Inosine (I)	+1.50	0.00	120
Cytidine (C)	+1.60	+0.10	47
Thymidine (T)	+1.70	+0.20	20

Table 3.1 – Reduction potential values for nucleosides in water and the difference compared to excited-state 2AP (2AP*). The lifetime shown is the fast decay component of transient absorption measurements of solutions of 2AP mixed with nucleotides. Reduction potentials from Steenken and Jovanovic[8] for the natural bases and Kelley and Barton[9] for 2AP* and I. Lifetimes are taken from Fiebig, Wan, and Zewail.[6]

Redox potential calculations performed by Crespo-Hernandez *et al.* corroborate the above findings; although, these calculations were performed in ACN for oxidation potentials and DMF for reduction potentials rather than water.[10]

A more recent study by Narayanan *et al.*[11] measured the oxidation potential of 2AP* to be −2.02 V in ACN and the reduction potential of 2AP* to be +1.44 V in DMF. From calculations of the free energy of charge transfer with nucleotides they predicted that, in water, rGMP and rAMP are oxidised by 2AP* but dTMP is reduced. The occurrence of charge transfer in rCMP was not conclusively determined but could potentially occur *via* either oxidation or reduction. It should be noted that the conclusions were based on the calculations of Crespo-Hernandez *et al.*[10] and experiments of Seidel, Schulz, and Sauer[7] and also required an estimate of the free energy decrease in going from organic to aqueous solution. The predictions of charge

transfer character are therefore based on a number of assumptions that may not be valid. For instance, absolute values from calculations are often dependent on the level of theory used and often only comparison of relative values is deemed to be appropriate. The use of an experimental value in combination with calculated values therefore seems somewhat questionable. Indeed, the findings are at odds with the conclusions of Crespo-Hernandez *et al*., despite the fact their calculations are used in the interpretation.

Despite the discrepancies, the findings of all these studies show that the intrinsic redox potential of the base neighbouring 2AP has a significant effect on the observed fluorescence behaviour.

Somsen *et al.* used time-resolved fluorescence measurements to investigate the temperature dependence of 2AP fluorescence in dinucleotides and used a four-exponential model to fit all fluorescence decay curves measured.[12] A summary of the fit parameters obtained at 20°C is shown in Table 3.2 (top row for each dinucleotide). To explain the decay parameters they invoke a model where there are three distinct ground-states but one of these states has two associated excited-states that are in equilibrium with each other. These linked excited states are associated with the A_1 and A_3 components. They proposed that transient unstacking occurs in the excited state (not the ground state) based on thermodynamic arguments but it is not clear why this could not be a ground-state phenomenon. They propose that their model is a variant on a gating mechanism used to explain charge transfer dynamics in DNA.[13,14]

Sample	Lifetimes /ns				A-Factors				$\langle\tau\rangle$ /ns
	τ_1	τ_2	τ_3	τ_4	A_1	A_2	A_3	A_4	
2AP-A	0.03	0.59	2.10	8.30	0.33	0.41	0.23	0.03	0.98
	-	0.61	2.00	9.60	-	0.64	0.33	0.03	1.34
2AP-C	0.04	0.45	2.50	8.00	0.48	0.14	0.36	0.02	1.14
	0.16	0.78	2.40	9.20	0.17	0.14	0.68	0.01	1.86
2AP-G	0.03	0.33	1.80	8.80	0.78	0.07	0.12	0.02	0.44
	0.06	0.63	1.80	10.30	0.47	0.14	0.33	0.06	1.33
2AP-T	0.03	0.38	2.40	8.20	0.60	0.10	0.29	0.01	0.84
	0.07	0.65	2.30	10.20	0.27	0.11	0.59	0.03	1.75
2AP-I	0.02	0.80	3.10	8.40	0.35	0.20	0.39	0.06	1.87
	-	0.83	3.10	8.50	-	0.33	0.61	0.06	2.67

Table 3.2 – Comparison of lifetime parameters obtained for 2AP-containing dinucleotides. Taken from Somsen *et al.*[12] (top row) and Wu[15] (bottom row).

Wu also studied 2AP-N dinucleotides in aqueous solution (as with Somsen *et al.*, phosphate buffered saline was used to maintain pH) and fit parameters are shown in Table 3.2 (bottom row for each dinucleotide).[15] The lifetime parameters obtained by Wu are generally quite consistent with those of Somsen *et al.*; lifetimes roughly fall into the same four distinct ranges that are typical for 2AP in nucleic acids and A-factors share fairly similar occupancies. The most obvious difference between the two sets of parameters is the model used to describe 2AP-A and 2AP-I. Wu found that only a three-component model was required to adequately fit the data; specifically, the shortest lifetime is absent. This discrepancy can be rationalised by considering the relative charge transfer efficiency of the bases. As highlighted above, Fiebig, Wan, and Zewail showed that the rapid component in complexes of 2AP and nucleotides was slowest with adenosine.[6] The lifetime measured (350 ps) was an order-of-magnitude greater than the other natural bases and is comparable to the typical lifetime of the second fastest component (~600 ps) measured by time-correlated single photon counting (TCSPC). Therefore, although two different relaxation pathways might be available, the limited resolution offered by the experimental method means that a single component is adequate during fitting; consequently, only three components are required to fit the fluorescence decay of the 2AP-A and 2AP-I dinucleotides. It is possible to argue the case for using either the three- or four-exponential model. Usually it is best to use the simplest model that adequately describes the system because there is no reason to assume additional complexity and over-fitting can lead to an unphysical interpretation; however, comparison between dinucleotide parameter sets is made far more intuitive if the same model is used for all data sets. It also seems unlikely that the component has been completely lost and higher resolution experiments might be able to separate the two components.

Nakabayashi, Islam, and Ohta investigated the fluorescence properties of flavin adenine dinucleotide (FAD) in various alcohol-water mixtures.[16] FAD is an intrinsically fluorescent biomolecule commonly found in cellular environments and its fluorescence character is sensitive to its conformational state. In particular, fluorescence lifetimes vary greatly depending on whether or not the molecule is in a stacked (ps lifetime) or open conformation (ns lifetime); this has similarities to 2AP-containing dinucleotides. The structure of FAD is shown in Figure 3.2.

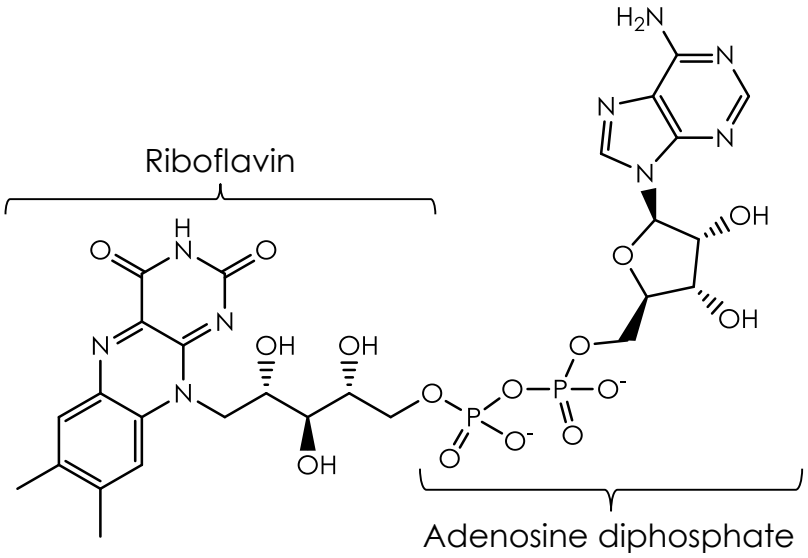

Figure 3.2 – Molecular structure of flavin adenine dinucleotide (FAD). The two moieties which given the molecule its name are highlighted. As it contains the sugar alcohol ribitol (rather than ribose), riboflavin is not technically a nucleotide. The term dinucleotide is therefore a misnomer.

Nakabayashi, Islam, and Ohta showed that the fluorescence lifetime of FAD was heavily influenced by the polarity of the solvent mixture. Increasing the alcohol content was found to increase the occupancy of the longest lifetime component. The change was attributed to an increase in the population of the open conformation at the expense of the stacked conformation. As alcohol content increases (solvent polarity decreases) it can be predicted that the open dinucleotide conformation will become more stable. This can be rationalised considering the fact that hydrophobic effects, which cause the bases to stack, will decrease as alcohol content increases. The same argument holds for nucleic acid dinucleotides and is consistent with the fact that the melting temperature of DNA decreases as alcohol concentration is increased[17,18] (although DNA melting concerns base-pairing, rather than base-stacking, the effect is essentially the same; the bases are more stable in isolation when polarity decreases). It also agrees with findings from the work presented in Chapter 5 that show that the energy difference between stacked and open dinucleotides is smaller in less polar solvents; implying the conformational equilibrium will be shifted towards the open state.

Given the obvious similarity between FAD and nucleic acid dinucleotides it seemed pertinent to use the same approach as Nakabayashi, Islam, and Ohta to gain insight into the conformational dynamics of 2AP-containing dinucleotides. The remainder of this chapter will describe the result of following this strategy.

3.2 Materials and Methods

3.2.1 Sample Preparation

Unless otherwise stated all chemicals were sourced from Sigma-Aldrich. 2AP and 2AP-riboside (r2AP, Santa Cruz Biotechnology) were dissolved in deionised water, ethanol, or glycerol to produce stock solutions (~15 μM). Samples were then prepared by mixing appropriate masses of each stock together.

2AP-containing dinucleotides (2AP-N, N = A, C, G, T, or I; ATDBio) were dissolved in deionised water to produce stock solutions of about 100-140 μM for 2AP-A, 2AP-C, 2AP-G, and 2AP-T and 200 μM for 2AP-I. Appropriate masses of alcohol, water, and stock were then combined to produce sample mixtures. Sample concentration (~10-20 μM) was determined by ultraviolet-visible (UV-Vis) spectroscopy using the maximum of 2AP absorbance (6,000 M^{-1} cm^{-1} at 305 nm).[19–21] Concentration could also be estimated by using the following extinction coefficients at 260 nm:[20,22] adenosine (dAMP), 15,060 M^{-1} cm^{-1}; cytidine (dCMP), 7,100 M^{-1} cm^{-1}; guanosine (dGMP), 12,180 M^{-1} cm^{-1}; thymidine (dTMP), 8,560 M^{-1} cm^{-1}; and 2AP, 1,700 M^{-1} cm^{-1}.

The dielectric coefficient, ε, and empirical polarity parameter, $E_T(30)$, of the pure solvents used in this investigation are shown in Table 3.3.

Solvent	ε	$E_T(30)$ /kcal mol^{-1}
Water	80.37	63.10
Glycerol	41.14	57.00
Ethanol	25.00	51.90
DMSO	45.00	45.10
1,4-Dioxane	2.25	36.00

Table 3.3 – Summary of pure solvent properties. The dielectric coefficients, ε, at 20°C were taken from Åkerlöf[23] for water, glycerol and ethanol and Evans *et al.*[24] for DMSO and 1,4-dioxane. All $E_T(30)$ values were taken from Reichardt.[25]

Characteristics of the alcohol-water mixtures used in this investigation are shown in Table 3.4. Solvent mixture labels represent the alcohol type (W, water; G; glycerol; and E, ethanol) and content % (w/w).

Solvent Mixture	Alcohol Content /% (w/w)	/Mole Fraction	ε
Water			
W100	0	0.00	80.37
Glycerol Series			
G20	20	0.05	74.72
G30	30	0.08	71.77
G50	50	0.16	65.63
G60	60	0.23	62.03
G70	70	0.31	57.06
G100	100	1.00	41.14
Ethanol Series			
E30	30	0.14	62.63
E50	50	0.28	50.38
E100	100	1.00	25.00

Table 3.4 – Summary of alcohol-water mixture properties. Dielectric coefficients, ε, at 20°C were taken from Åkerlöf.[23] An error of ≤ 5% is estimated for alcohol content in the mixtures.

3.2.2 Steady-State Measurements

Samples were measured in a UV quartz cell (Starna) with a 3 mm path length. Absorption measurements were carried out on a Cary50 Bio UV-Vis spectrometer (Varian). Triplicate spectra were recorded from 500 nm to 200 nm in 0.5 nm steps for each sample. Water was used for a reference. Excitation and emission spectra were recorded on a Cary Eclipse Fluorescence Spectrometer (Varian). Triplicate emission spectra were recorded between 320 nm and 500 nm in 0.5 nm steps. Samples were excited at 300 nm. Excitation and emission slits were set for 5 nm bandwidth. Excitation spectra were recorded between 200 nm and 350 nm using the same parameters as above. Emission was collected at 370 nm. All measurements were made at room temperature (22°C). Data were analysed and plotted using a combination of MATLAB[26] and Microsoft Excel (2010).[27]

3.2.3 Time-Correlated Single Photon Counting (TCSPC)

Samples were measured in a UV quartz cell (Starna) with a 3 mm path length. Fluorescence decay curves were recorded using a TCSPC Box equipped with Becker and Hickl photon counting electronics. The excitation source was a tuneable, mode-locked Ti:Sapphire laser system (Coherent, MIRA 900f Ti:Sapphire laser pumped by a Coherent, 10W Verdi CW laser). The MIRA cavity length was extended with the

addition of a PulseSwitch cavity dumper unit (Coherent, APE PulseSwitch), thereby reducing the fundamental repetition rate of the laser to 54 MHz. The cavity dumper was used to reduce the pulse repetition rate to 5.4 MHz, producing pulses of approximately 200 fs duration. A harmonic generator system (Inrad Optics, Model 5-050) was used to triple the frequency of the source light. Fluorescence emission was detected orthogonal to the excitation beam through a polariser set at the magic angle (54.7°) with respect to the vertically polarised excitation. Using the magic angle eliminates possible biases caused by any anisotropy present in the fluorescence emission. A lens was used to collect emission before it was spectrally filtered using a monochromator (JY H20). Photons were detected using a cooled microchannel plate detector (Eldy, EM132-1). All experiments were performed at an excitation wavelength of 300 nm. A 320 nm long pass filter (Schott Glass, Newport) was used to block scattered light during fluorescence decay measurements. Fluorescence decay curves were recorded on a time scale of approximately 40 ns, binned into 2,048 channels, to a total of at least 10,000 counts in the peak channel. Decays were measured at an emission wavelength of 370 nm. All measurements were made at room temperature (22°C). The instrument response function (IRF) was recorded using scattered light from Ludox solution at 300 nm. An example of an IRF recorded during this work is shown in Figure 3.3. The IRF can be considered to be the intensity profile that is detected from an excitation pulse without the presence of any sample.

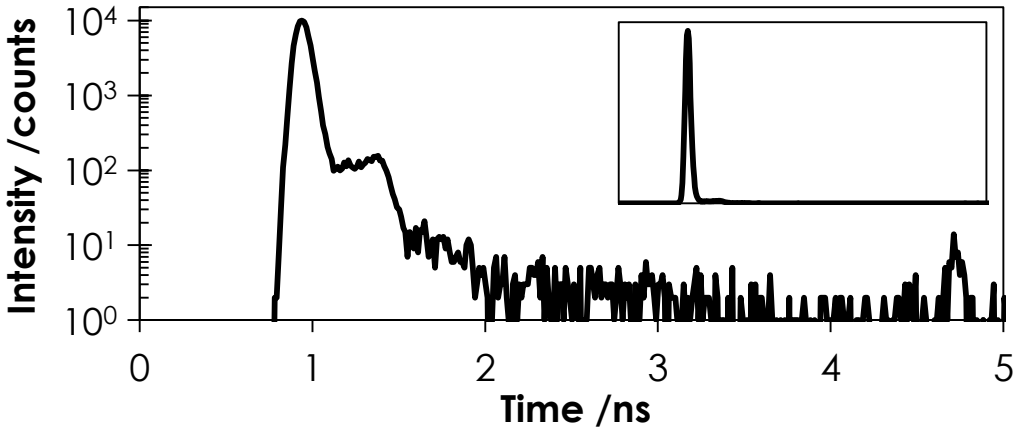

Figure 3.3 – A typical instrument response function (IRF) measured for time-correlated single photon counting (TCSPC) experiments during this work. The IRF was measured by scattering 300 nm light from a suspension of colloidal silica (Ludox, Sigma-Aldrich) at 22°C. The inset shows the same data on a linear intensity scale and highlights dominant contribution of the main peak compared to the 'afterpulse' structure.

Although excitation pulses had a temporal width of around 200 fs, the full-width at half-maximum of the main IRF peak is around 90 ps. The most significant cause of broadening in the IRF is the transit time spread of electrons passing through the microchannel-plate photomultiplier tube (MCP-PMT) detector. Transit time spread is a manifestation of the variability in the distance the electrons travel as they are amplified within the MCP-PMT. Other factors that affect the IRF include the wavelength of light detected and the settings that are used for any electronic components in the detector system. As well as broadening, the low intensity 'afterpulse' that follows the main peak must also be accounted for during analysis of the measured fluorescence decays to achieve reliable fits.

It should be noted that, due to time constraints, only a single decay was typically recorded for each sample in this study. Along with the fairly low number of counts in the peak channel, this means that the accuracy of the results presented, particularly for the multi-exponential dinucleotide decays, should be treated with caution. This being said, as decays were analysed as part of a series, the results still offer qualitative insight into trends in the photophysical behaviour of 2AP in isolation and within nucleic acids.

Decay curves were analysed by using an iterative reconvolution method in the FAST software package.[28] A multi-exponential decay was assumed;

$$I(t) = \sum_{i=1}^{n} A_i \cdot exp\left(-\frac{t}{\tau_i}\right), \qquad 3.1$$

where A_i and τ_i are the fractional amplitude and the fluorescence lifetime, respectively, of the i^{th} decay component and *n* is the total number of exponentials used in the fit. For comparison between samples, A-factors were normalised such that;

$$\sum_{i=1}^{n} A_i = 1.0. \qquad 3.2$$

The fit quality was judged on the basis of the reduced chi-squared statistic, χ_R^2, (which will be discussed in more detail in Chapter 4) and the randomness of the residuals. Typically a χ_R^2 value of less than 1.1 indicated an acceptable fit in this study.

An example of a rejected fit and an accepted fit is illustrated below. A bi-exponential decay was simulated and fitted with a mono-exponential function

(Figure 3.4) and a bi-exponential function (Figure 3.5); fit parameters are given in Table 3.5.

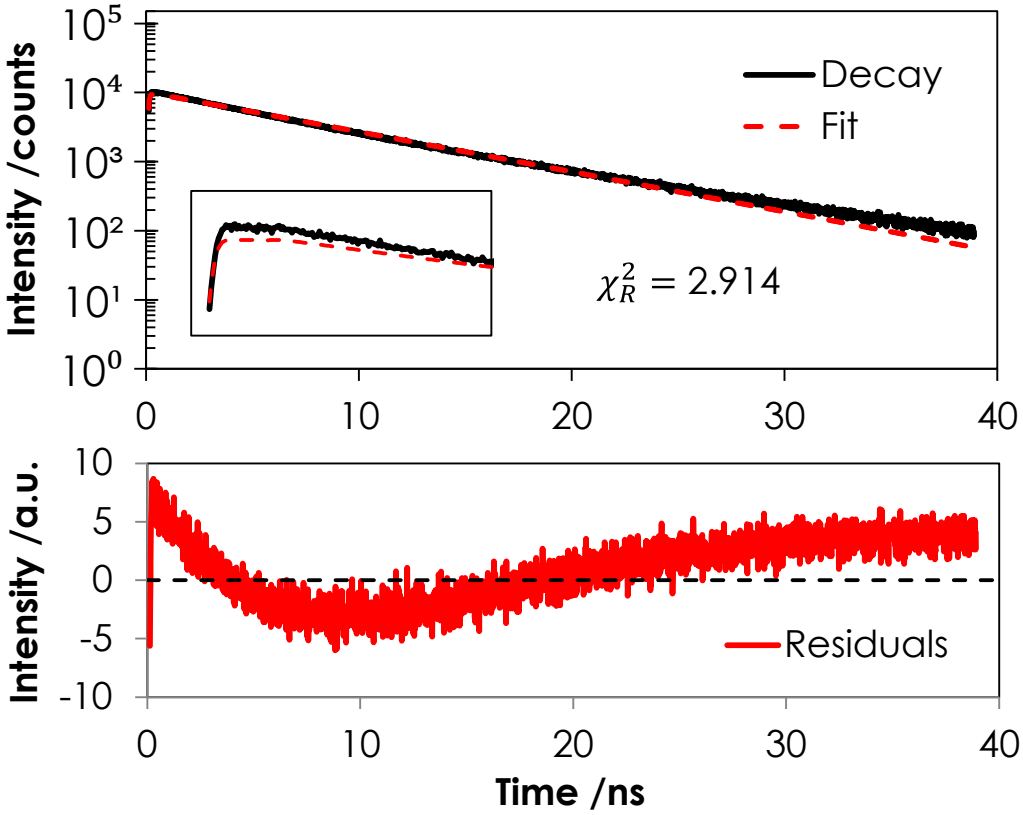

Figure 3.4 – An example of a rejected fluorescence decay fit. (Top) Mono-exponential fit of a simulated, bi-exponential decay. The inset shows early times (0 to 2 ns) on a linear intensity scale. (Bottom) Weighted residuals of the fit compared to the decay.

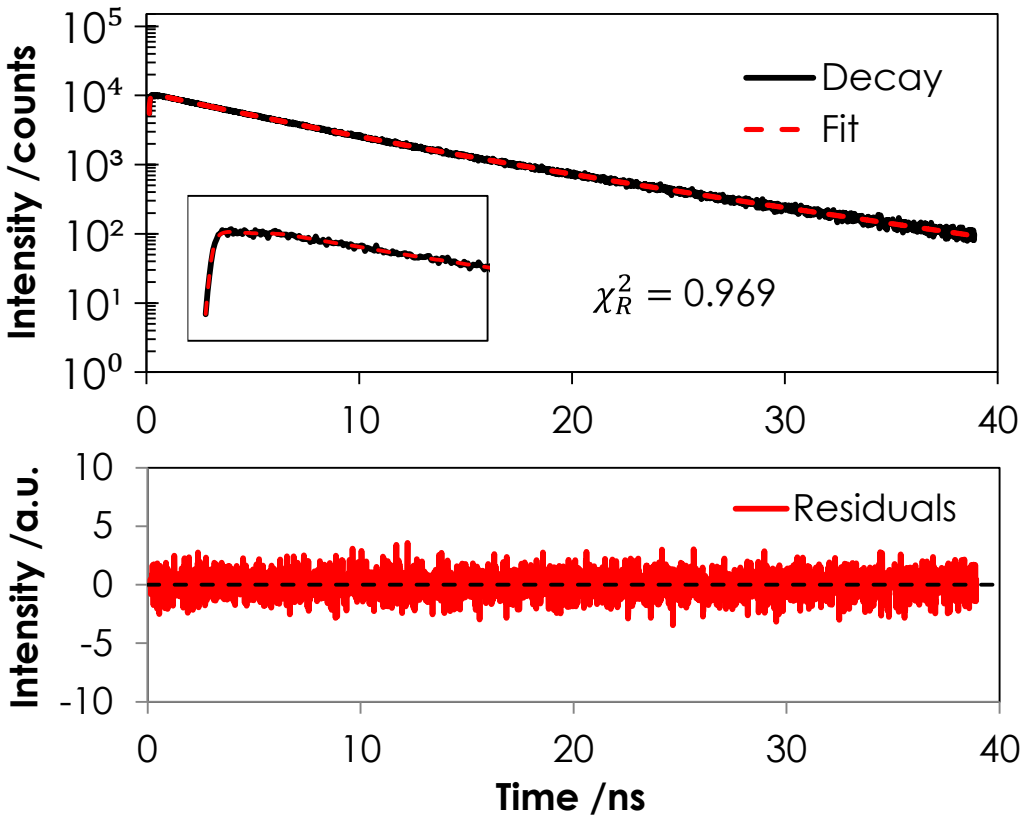

Figure 3.5 – An example of an acceptable fluorescence decay fit. (Top) Bi-exponential fit of a simulated, bi-exponential decay. The inset shows early times (0 to 2 ns) on a linear intensity scale. (Bottom) Weighted residuals of the fit compared to the decay.

Decay	Lifetime /ns		A-Factor		χ_R^2
	τ_1	τ_2	A_1	A_2	
Simulated	5.00	10.00	0.60	0.40	-
Mono-exponential	7.50	-	1.00	-	2.914
Bi-exponential	5.00	9.97	0.59	0.41	0.969

Table 3.5 – Mono-exponential and bi-exponential fit parameters obtained for a simulated bi-exponential decay.

The large χ_R^2 value and the non-random distribution of the residuals indicate that a mono-exponential fit does not adequately describe the data and so would be rejected as a possible model of the system. In contrast, the near-unity χ_R^2 value and the random distribution of the residuals indicate that a bi-exponential fit adequately describes the data and so would be accepted as a possible model of the system. As the bi-exponential

model provides an acceptable fit it would generally be considered inappropriate to fit the decay with a more powerful model, such as a three-exponential function.

It should be noted that, regardless of the fit quality of other models, a four-exponential model was indiscriminately used for the fluorescence decay of the 2AP-containing dinucleotides in this chapter. This was done to allow a more intuitive comparison of the results but, as discussed later, may have introduced artefacts due to over-fitting in some cases.

Amplitude-weighted lifetimes, $\langle\tau\rangle$, give an approximate measure of the relative emission quantum yield for analogous systems and were calculated *via* Equation 3.3;

$$\langle\tau\rangle = \sum_{i=1}^{n} A_i \cdot \tau_i. \qquad 3.3$$

3.3 Results and Discussion

3.3.1 Steady-State Spectroscopy of 2AP

3.3.1.1 2AP in Water

Steady-state spectra collected for 2AP in water are shown in Figure 3.6. Absorption and excitation spectra are almost identical and show a maximum around 305 nm. The emission spectrum shows a single peak with a maximum around 370 nm.

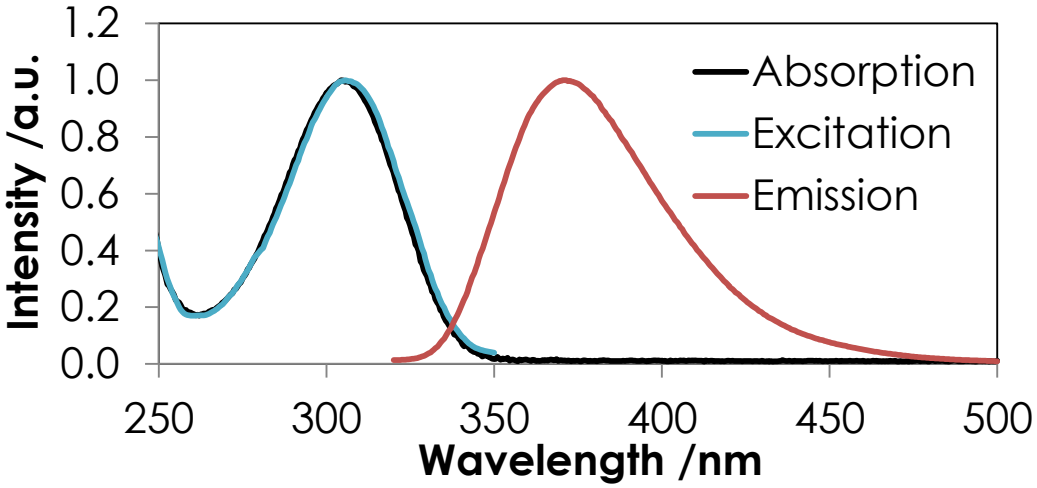

Figure 3.6 – Steady-state spectra for 2AP in water. 2AP concentration was ~15 µM. Emission was collected at 370 nm for the excitation spectrum. The emission spectrum was recorded with 300 nm excitation.

3.3.1.2 2AP in Alcohol-Water Mixtures

Absorption spectra collected for 2AP in various ethanol-water mixtures are shown in Figure 3.7. The absorption peak is red-shifted as alcohol content increases.

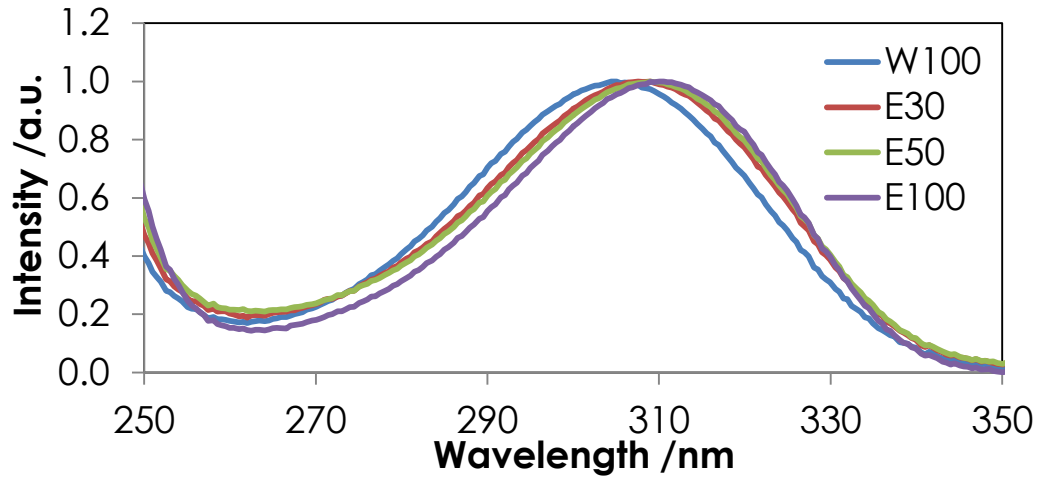

Figure 3.7 – Steady-state absorption spectra for 2AP in ethanol-water mixtures. 2AP concentration was ~15 µM for all samples. The numbers in the legend entries show the alcohol content % (w/w); W100 is pure water.

Excitation and emission spectra collected for 2AP in various ethanol-water mixtures are shown in Figure 3.8. The excitation and emission peaks are red- and blue-shifted, respectively, as alcohol content increases, resulting in a decrease in Stokes shift with increasing alcohol content.

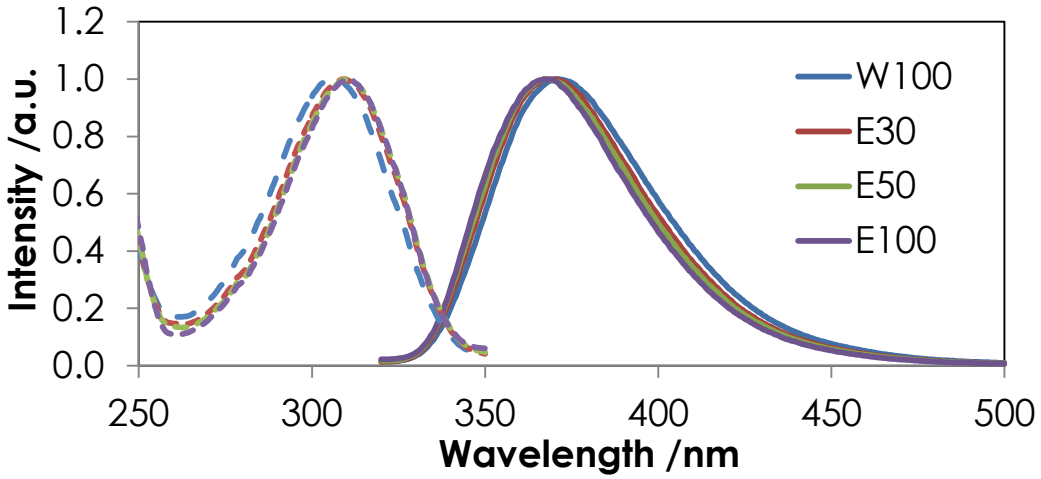

Figure 3.8 – Steady-state excitation (dashed lines) and emission (solid lines) spectra for 2AP in ethanol-water mixtures. 2AP concentration was ~15 µM for all samples. Emission was collected at 370 nm for the excitation spectrum. The emission spectrum was recorded with 300 nm excitation. The numbers in the legend entries show the alcohol content % (w/w).

Absorption spectra collected for 2AP in various glycerol-water mixtures are shown in Figure 3.9. The absorption peak is red-shifted as alcohol content increases.

The increase in optical density at short wavelengths is due to absorption by glycerol and is likely a consequence of using pure water as a reference for all measurements.

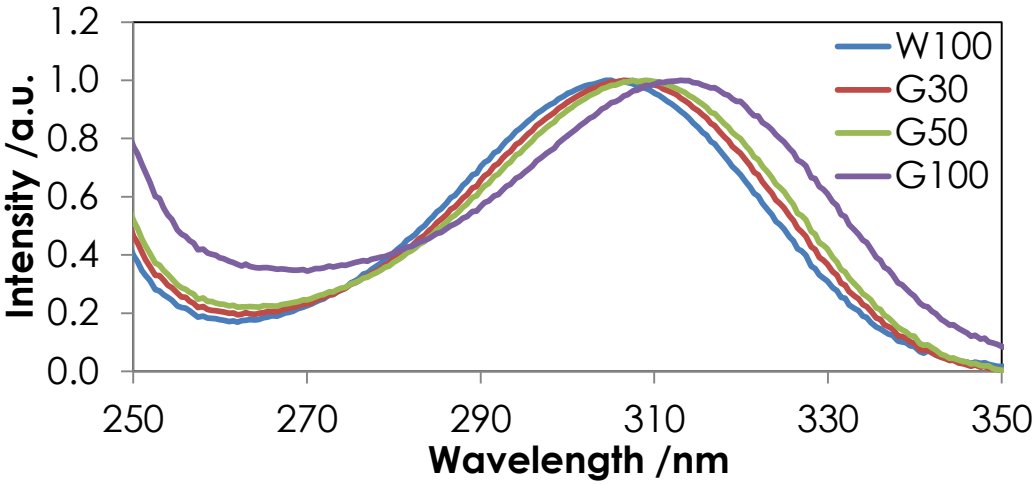

Figure 3.9 – Steady-state absorption spectra for 2AP in glycerol-water mixtures. 2AP concentration was ~15 µM for all samples. The numbers in the legend entries show the alcohol content % (w/w).

Excitation and emission spectra collected for 2AP in various glycerol-water mixtures are shown in Figure 3.10. The excitation peak is red-shifted as alcohol content increases but the emission peak remains unchanged, consistent with Evans *et al.*[24] The increase observed in the absorption spectra at short wavelengths is not replicated in the excitation spectra, which confirms that the effect is not due to changes in 2AP absorption.

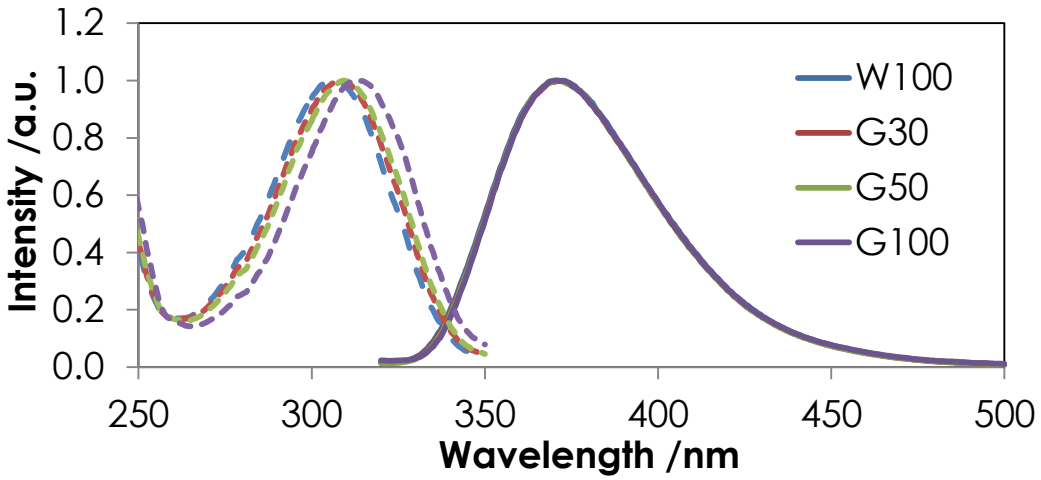

Figure 3.10 – Steady-state excitation (dashed lines) and emission (solid lines) spectra for 2AP in glycerol-water mixtures. 2AP concentration was ~15 µM for all samples. Emission was collected at 370 nm for the excitation spectrum. The emission spectrum was recorded with 300 nm excitation. The numbers in the legend entries show the alcohol content % (w/w).

The red-shift in the excitation spectra observed for both alcohols can be understood by considering the relative stabilisation effect of each solvent on the ground state and the Franck-Condon excited state. A schematic of the effect of solvent relaxation on 2AP fluorescence is shown in Figure 3.11.

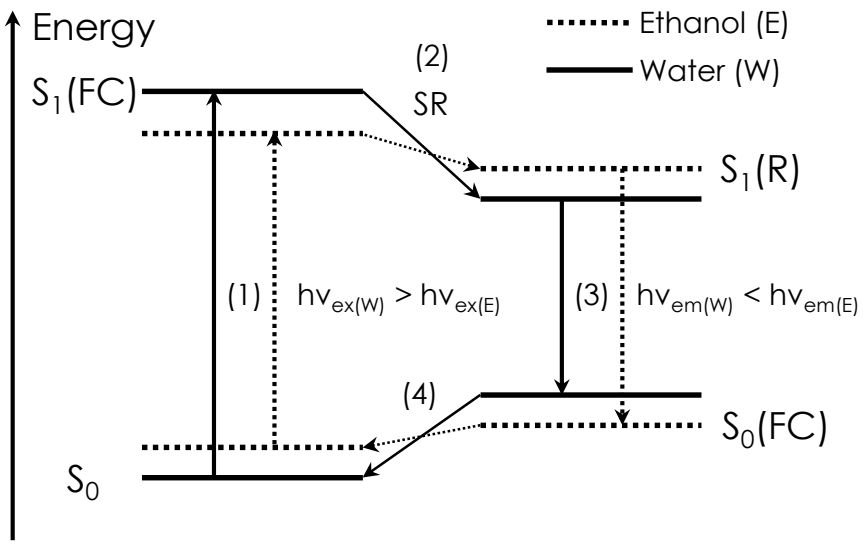

Figure 3.11 – Schematic of solvent relaxation (SR) effects on 2AP in water (W, solid lines) and ethanol (E, dotted lines). Excitation (1) and emission (3) follow vertical transitions and produce Franck-Condon (FC) states. Solvent relaxation (2 and 4) stabilises the FC states by compensating for changes in the dipole moment of 2AP. Higher polarity solvents can stabilise the dipole moment of 2AP more effectively; however, in the Franck-Condon states, before solvent relaxation has occurred, they cause more of a destabilisation effect. Therefore 2AP excitation is red-shifted and emission is blue-shifted in ethanol compared to water. Figure based on Lackowicz.[29]

The lower polarity of the alcohols compared to water means that any change in charge distribution (dipole moment) requires a less extensive reorganisation of solvent. In effect, solvent orientation in alcohol is less important than in water and so the ground and Franck-Condon solvated states are closer in energy; resulting in a red-shifted excitation.

Emission generally occurs after solvent relaxation has occurred and so the solvent-dependent emission spectral shift can be explained by considering the difference between the relaxed excited state and the unrelaxed ground state. The blue-shift in emission observed for ethanol is a result of poorer stabilisation of the excited state of 2AP compared to water. In addition, the unrelaxed ground state is likely to be lower in energy in ethanol than water for the same reason that was described for the Franck-Condon excited state; in water, the relaxed excited state solvation shell is further from its optimal ground-state orientation.

The overlap of the emission peak of 2AP in glycerol and water means that the energy difference between the relaxed excited state and the unrelaxed ground state is the same, within the spectral resolution of these measurements, regardless of the

solvent composition. That is to say, any solvent dependent change in the energy of the relaxed excited sate is compensated by an equivalent change in the unrelaxed ground-state.

3.3.2 Steady-State Spectroscopy of 2AP-N Dinucleotides

Steady-state absorption spectra for 2AP, the natural deoxynucleoside monophosphates (dAMP, dCMP, dGMP, and dTMP), and 2AP-N dinucleotides (where N = A, C, G, T, or I) in water are shown in Figure 3.12. Importantly, absorption by the natural base is negligible at around 300 nm meaning that 2AP can be selectively excited when incorporated in nucleic acid constructs.

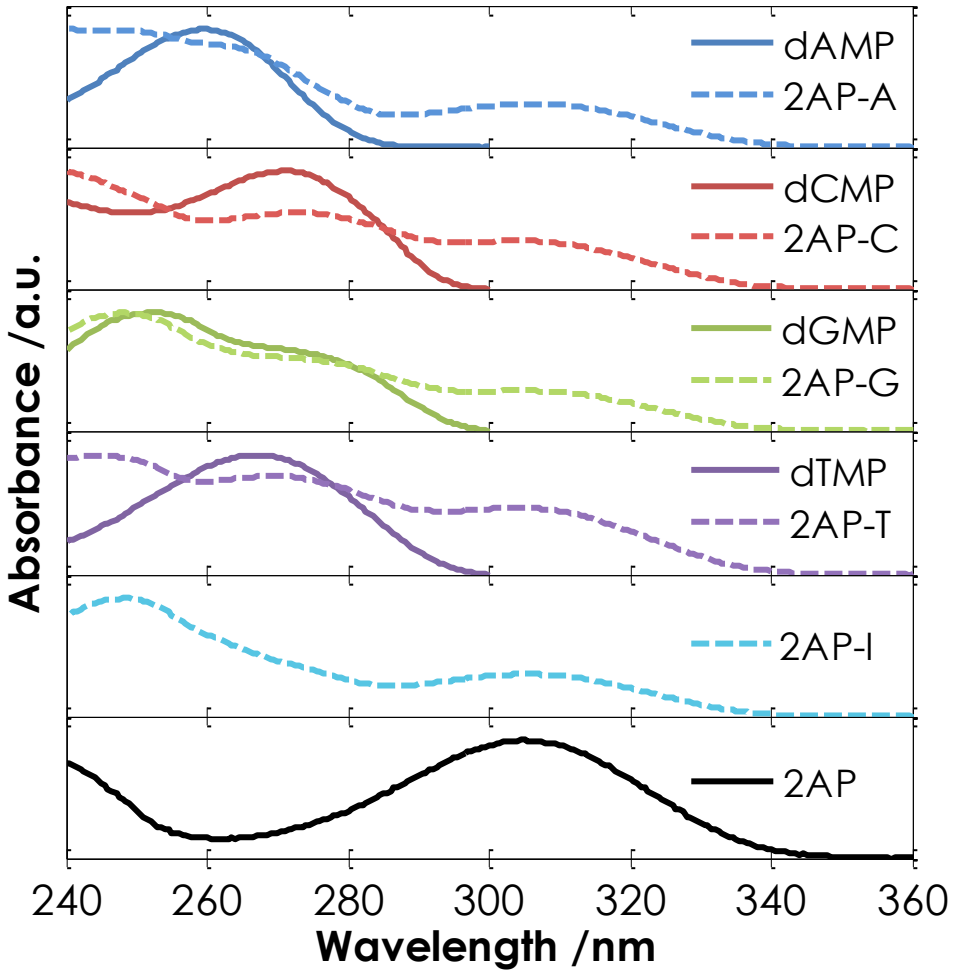

Figure 3.12 – Steady-state absorption spectra for 2AP, the natural deoxynucleoside monophosphates, and 2AP-N dinucleotides in water. Absorption spectra for the natural bases were taken from Cavaluzzi and Borer.[22] Absorption data for the nucleoside inosine was unavailable.

Excitation and emission spectra for 2AP and 2AP-containing dinucleotides in water are shown in Figure 3.13.

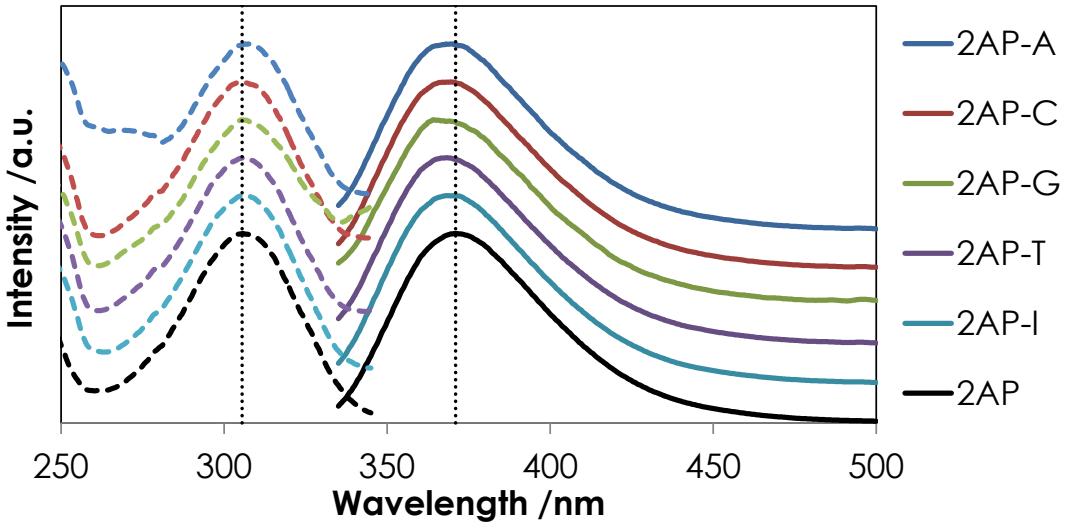

Figure 3.13 – Fluorescence spectra collected for 2AP-containing dinucleotides in water. Dashed lines show excitation spectra while solid lines show emission spectra. Sample concentration was ~10-20 μM. Spectra have been normalised and shifted vertically for clarity. The black dotted lines are shown to highlight the peak position of the 2AP spectra. Emission was collected at 370 nm for the excitation spectra. The emission spectra were recorded with 300 nm excitation.

The observed spectra are very similar for all dinucleotides. Excitation and emission spectra are red- and blue-shifted, respectively, for the dinucleotide compared to free 2AP. Similar to the effect of adding alcohol, solvent stabilisation is reduced for 2AP-dinucleotides as the neighbouring base acts as a (non-polar) barrier to solvent interactions. Although it appears that the 2AP-G excitation spectrum increases in intensity at around 340 nm this is likely an artefact of the measurement and simply due to scattered light. It is possible that the effect is most evident in 2AP-G because this dinucleotide has the lowest emission quantum yield and is therefore more susceptible to noise.[5,15] It is worth mentioning that in alcohol mixtures, where 2AP-G fluorescence intensity is greater, the anomalous intensity increase was severely diminished. A striking feature in the 2AP-A excitation spectra is the enhanced intensity at around 260 nm compared to the other bases. This correlates with the maximum absorption of adenine, as shown in Figure 3.12, and can be attributed to energy transfer from the natural base to 2AP. Energy transfer efficiency has been found to be much greater for adenine that the other nucleobases and so it is unsurprising that 2AP-A is the only dinucleotide to exhibit this effect.[30,31]

Figure 3.14 shows the effect on the dinucleotide fluorescence spectra of introducing alcohol into the solvent. Only 2AP-A and 2AP-C spectra are shown as the latter is representative of all other dinucleotides.

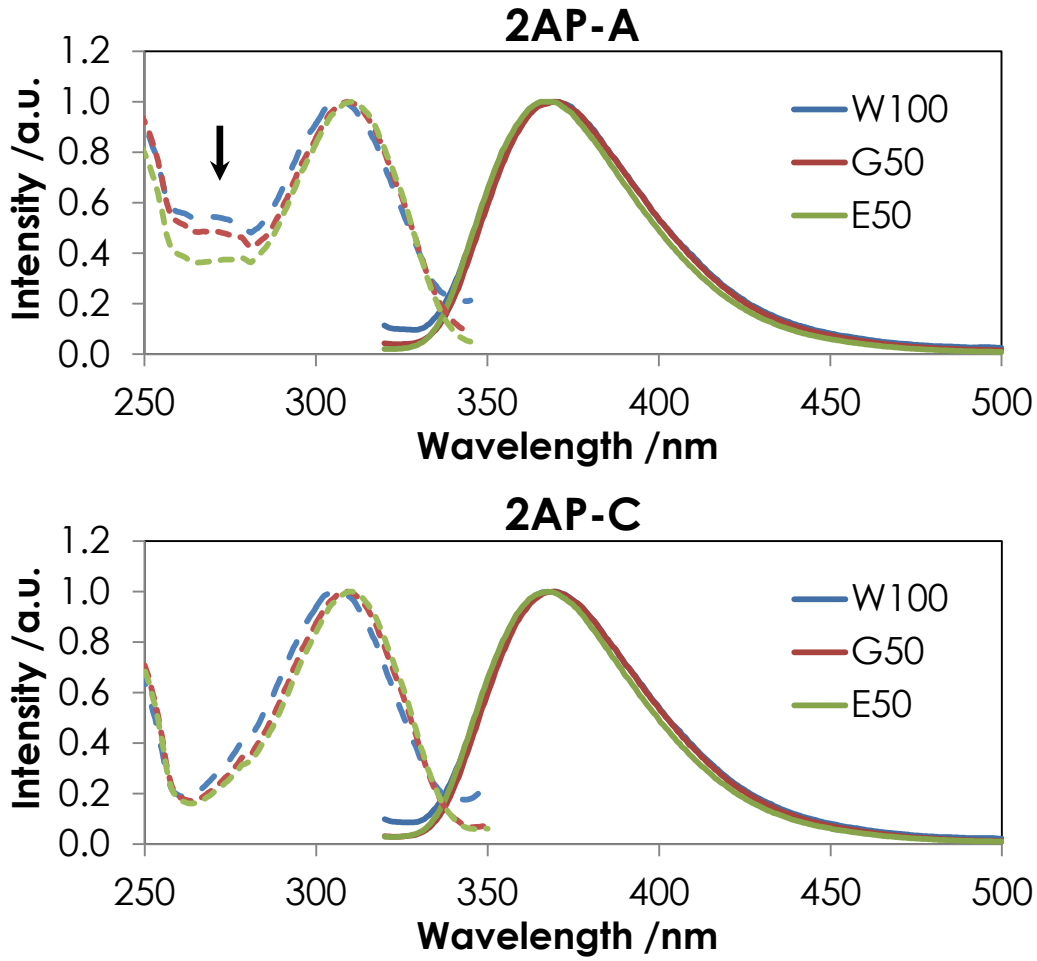

Figure 3.14 – Comparison of 2AP-A (top) and 2AP-C (bottom) fluorescence spectra in water and glycerol-water (G50) and ethanol-water (E50) mixtures. Dashed and solid lines show excitation and emission spectra, respectively. Sample concentration was ~10-20 μM in all cases. The black arrow highlights the reduction in intensity at around 260 nm for 2AP-A when alcohol is present; the reduction is not observed in 2AP-C. Emission was collected at 370 nm for excitation spectra. Emission spectra were recorded with 300 nm excitation.

2AP-A and 2AP-C spectra show similar changes to that observed for free 2AP; excitation and emission spectra are red- and blue-shifted, respectively, for the ethanol-water mixture but only the excitation spectrum shifts (again, to the red) in the glycerol-water mixture. As mentioned above, the neighbouring base will shield 2AP from solvent and so it is unsurprising that the relative shifts appear to be smaller than

those observed for free 2AP. Interestingly, the excitation spectrum of 2AP-A shows a reduction in intensity at around 260 nm, which implies that energy transfer efficiency is reduced in the alcohol-water mixtures. This is consistent with the idea that an open dinucleotide conformation, where bases are unstacked and separated from each other, will become more favourable as solvent polarity decreases (alcohol content increases). A similar change in energy transfer efficiency was observed by Evans *et al.* when changing the temperature of a duplex that contained 2AP next to adenine.[24] Their results showed that energy transfer efficiency reduced as the temperature approached the duplex melting temperature; this can be explained by a loss of stacking efficiency as the duplex strands separate and is corroborated by the fact the excitation spectrum was observed to blue-shift with increasing temperature, which indicates 2AP became more solvent (water) accessible.

3.3.3 Time-Resolved Spectroscopy of 2AP

3.3.3.1 2AP in Water

Normalised fluorescence decay curves for 2AP, 2AP-riboside (r2AP) and 2AP-A in water are shown in Figure 3.15. Fluorescence decay curves for 2AP and r2AP are similar; both are mono-exponential and are characterised by a fairly long lifetime (~10 ns). The presence of the ribose group appears to have only minor influence on the fluorescence properties of 2AP; there is only a slight (~10%) decrease the decay lifetime. The 2AP deoxyribose sugar group in the dinucleotides was assumed to have similar effect to the ribose here. The fluorescence decay curve of 2AP-A is very different; the decay is multi-exponential and attenuates much more rapidly, which indicates that when 2AP is incorporated into a dinucleotide overall quenching is greater and the environment is much more heterogeneous. The decay of 2AP-A is shown here for comparison to free 2AP but will be discussed in more detail (along with other dinucleotides) in a later section.

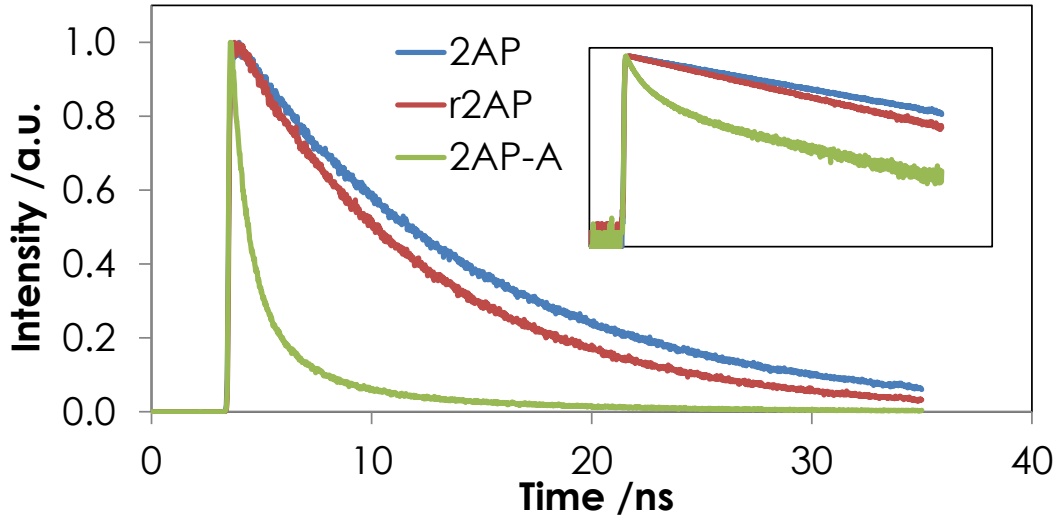

Figure 3.15 – Normalised fluorescence decay curves for 2AP, 2AP-riboside (r2AP), and dinucleotide 2AP-A in water. Sample concentration was ~10-20 μM in all cases. Excitation at 300 nm and emission collected at 370 nm. The inset shows the same decays on a logarithmic-scale.

3.3.3.2 2AP in Water and Other Pure Solvents

The fluorescence lifetime measured for 2AP in various solvents is shown in Figure 3.16. The lifetime values are plotted against the empirical polarity parameter of the solvent; specifically the $E_T(30)$ scale is used. Values for this scale are derived from the relative solvatochromatic shifts observed in each solvent for a standard dye which undergoes $\pi \rightarrow \pi^*$ absorption.[25] Although the linear model presented might be simplistic, it is clear that as polarity increases there is an increase in the fluorescence lifetime of 2AP; a similar observation was made from transient absorption measurements.[6] The trend is also consistent with the observations of Reichardt *et al.*;[1] the barrier to triplet formation is reduced as solvent polarity decreases and so the rate of relaxation through this channel increases, which consequently reduces the fluorescence lifetime.

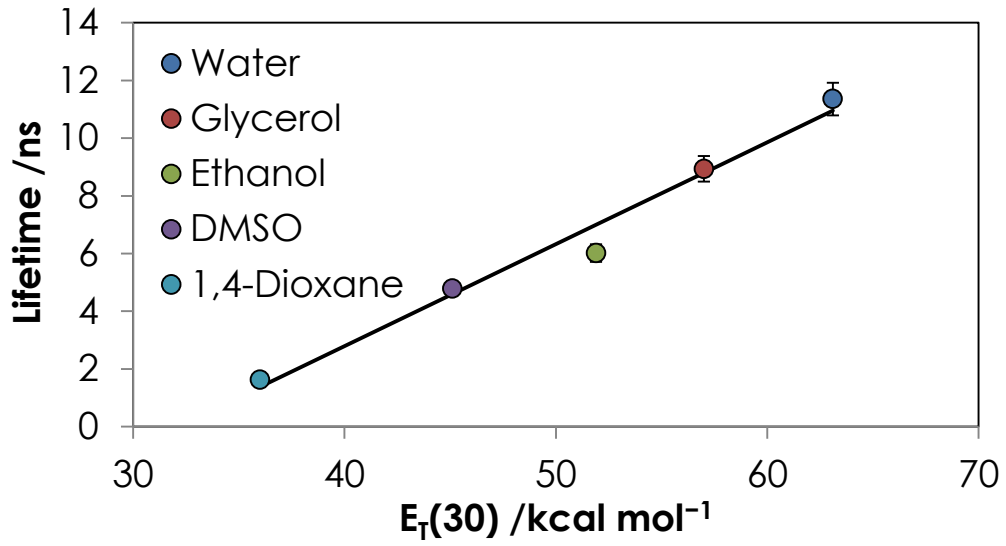

Figure 3.16 – Fluorescence lifetime of 2AP in various solvents. An estimated error of 5% is shown on values measured in this work (water, glycerol, and ethanol). The solid black line represents a linear fit to the data. $E_T(30)$ values were taken from Reichardt.[25] The lifetime values for DMSO and 1,4-dioxane were taken from Neely.[32]

3.3.3.3 2AP in Alcohol-Water Mixtures

Fluorescence lifetime parameters calculated for 2AP in various alcohol-water mixtures are summarised in Table 3.6.

Solvent Mixture	Alcohol Content /% (w/w)	Alcohol Content /Mole Fraction	ε	τ /ns
Water				
W100	0	0.00	80.37	11.35
Glycerol Series				
G20	20	0.05	74.72	11.18
G30	30	0.08	71.77	10.98
G50	50	0.16	65.63	10.49
G60	60	0.23	62.03	10.21
G70	70	0.31	57.06	9.92
G100	100	1.00	41.14	8.94
Ethanol Series				
E30	30	0.14	62.63	10.72
E50	50	0.28	50.38	10.04
E100	100	1.00	25.00	6.02

Table 3.6 – Fluorescence lifetimes measured for 2AP in various alcohol-water mixtures. Dielectric coefficients, ε, at 20°C were taken from Åkerlöf.[23] Alcohol content and fluorescence lifetimes, τ, are estimated to have an error of ≤ 5%.

The fluorescence lifetime of 2AP decreases with increasing alcohol content, which is consistent with the decrease in polarity (represented, roughly, by dielectric coefficient) expected when moving towards pure alcohol. It should be noted that samples with high alcohol content sometimes required additional components for adequate fitting; however, these components were minor and likely due to impurities which were susceptible to UV excitation. A full summary of the fits can be found in Appendix 3, §3.1. The value presented in Table 3.6 was the longest lifetime in each case and accounted for the vast majority of the emission; this lifetime was attributed to 2AP. The value obtained for pure ethanol (E100) is consistent with previous work[2] suggesting that the extraneous components do not extensively perturb the measured lifetime; however, this complication should be taken into consideration for the analysis of the fluorescence decays of 2AP-containing dinucleotides presented below in §3.3.4.

Figure 3.17 shows the dependence of the fluorescence lifetime on the alcohol content of the solvent.

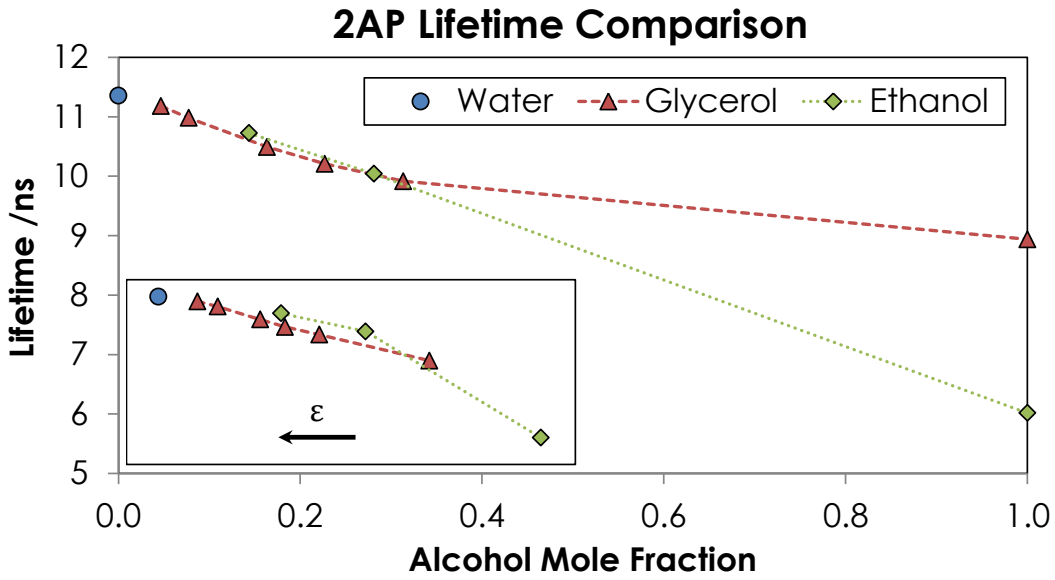

Figure 3.17 – Fluorescence lifetime of 2AP in various alcohol-water mixtures. 2AP concentration was ~15 μM in all cases. Dashed and dotted lines connecting data points have no physical meaning but are displayed to guide the eye. The insert shows the same lifetimes plotted against dielectric coefficient, ε, of the mixture. The direction of *increasing* ε is highlighted by the arrow. For clarity error bars are not shown but alcohol content and fluorescence lifetimes are estimated to have an error of ≤ 5%.

As is evident from Table 3.6, there is a general trend of decreasing lifetime with increasing alcohol content. The ethanol series shows an approximately linear change in

2AP lifetime between pure water and pure ethanol but the glycerol series is slightly more complicated; specifically, the 2AP lifetime in pure glycerol appears to be anomalously long when compared to the trend observed for the mixtures.

The previous section showed that solvent polarity plays a significant role in determining the 2AP lifetime in pure solvents. Plotting the lifetimes against dielectric coefficient (see insert of Figure 3.17) suggests that this is generally true for mixtures too; the observed lifetime decreases as polarity decreases. Interestingly, the glycerol series now shows linearity for all samples while the ethanol series does not (although this is based on few data points).

The properties of solvent mixtures are often more complicated than simply an amalgamation of the individual components and so it is hardly surprising that a single solvent parameter cannot account for the observed behaviour in all of the samples. One of the potential complications with mixtures is that solvent specific effects can cause heterogeneity in the system. Speculatively, it is possible that the local environment of 2AP is similar for the mixtures investigated, which all have quite low alcohol content. Water might dominate the solvation shell surrounding 2AP (even as alcohol content increases) because of more favourable solvent-solute interactions, such as H-bonding. The gradual decrease in 2AP lifetime observed might be explained by changes in bulk properties of the solvent mixture or by disruption of the water network surrounding 2AP. The latter proposal would be consistent with a recent study which suggested that H-bonding with a water molecule hindered rotation of the 2AP amino-group; a possible nonradiative relaxation mechanism.[33] Disrupting the amino-group H-bond would cause a decrease in lifetime by reactivating the relaxation pathway. If glycerol and ethanol cause a similar perturbation at equivalent (low) concentrations then this could explain why the lifetime observed appears to be independent of the alcohol used.

This being said, solvent viscosity should influence the observed lifetime of 2AP if mobility of the amino-group is a dominant factor. There is a 50-fold increase between G70 (18 cP) and G100 (906 cP) samples (for comparison, all other samples studied had a viscosity of only around 1-5 cP)[34,35] and so the lifetime of the 2AP in pure glycerol should be much longer than any mixture; however, this does not coincide with the experimental results as the lifetime is shorter (albeit perhaps higher than expected when plotting against mole fraction of alcohol). Additionally, the fact that the pure glycerol lifetime sits on the linear plot shown in Figure 3.16 suggests that the lifetime can essentially be explained by polarity effects and so viscosity cannot hold significant influence on this system (otherwise it would be an outlier).

While conformational dynamics of the amino-group might not be a dominant factor in determining the fluorescence lifetime of 2AP it is possible that the observed behaviour is still caused by disruption of the water network. There is strong evidence that suggests that the H-bonding ability of a pure solvent has influence on the observed fluorescence character of 2AP.[1]

It is clear that further investigation is required to make confident predictions about the precise causes of the fluorescence behaviour of 2AP in solution, particularly in mixtures.

3.3.4 Time-Resolved Spectroscopy of 2AP-N Dinucleotides

3.3.4.1 2AP-N Dinucleotides in Water

Figure 3.18 shows the fluorescence decay curves collected for 2AP-containing dinucleotides in water.

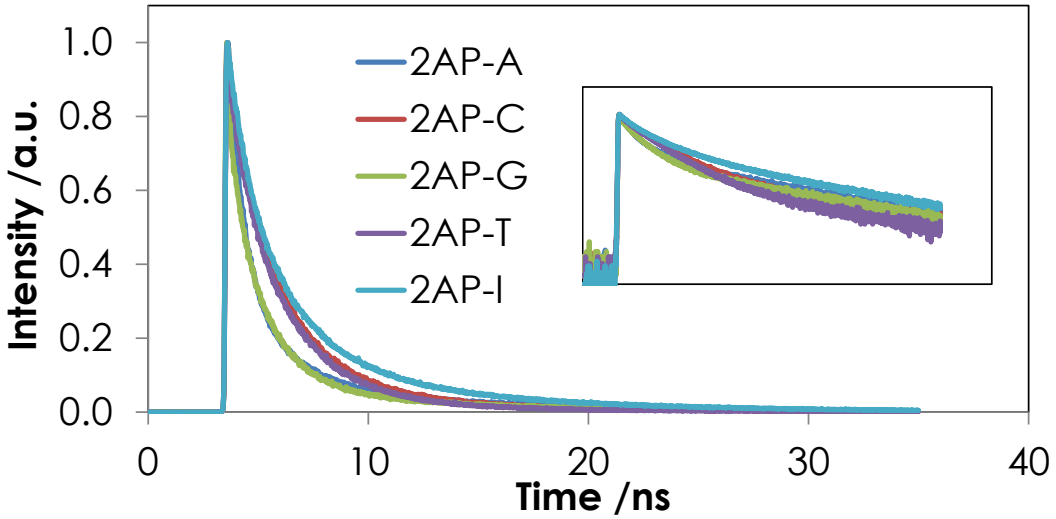

Figure 3.18 – Fluorescence decay curves for 2AP-containing dinucleotides in water. The inset shows the same decays on a logarithmic-scale. In all cases sample concentration was ~10-20 µM. Excitation at 300 nm and emission collected at 370 nm. Note that the decays of 2AP-A and 2AP-G are almost completely overlapped; similarly for 2AP-C and 2AP-T.

The fluorescence decay curve of all of the 2AP dinucleotides shows rapid attenuation in intensity and exhibits multi-exponential behaviour. Interestingly, the decays of dinucleotides containing proximal purine bases (A and G) appear to show distinct characteristics from those containing pyrimidine bases (C and T). 2AP-I fluorescence intensity decays less rapidly than the dinucleotides containing canonical nucleic acid bases. Associated fit parameters for these decays are given in Table 3.7.

W100	Lifetimes /ns				A-Factors				χ_R^2	$\langle\tau\rangle$ /ns
	τ_1	τ_2	τ_3	τ_4	A_1	A_2	A_3	A_4		
2AP-A	0.08	0.65	2.13	8.97	0.27	0.45	0.22	0.06	0.991	1.293
2AP-C	0.06	0.58	2.45	8.96	0.40	0.13	0.44	0.04	0.967	1.492
2AP-G	0.04	0.51	1.70	8.98	0.66	0.09	0.22	0.03	1.122	0.687
2AP-T	0.05	0.66	2.45	9.36	0.55	0.09	0.34	0.01	0.987	1.070
2AP-I	0.06	0.85	2.69	9.20	0.35	0.20	0.36	0.09	0.953	1.952

Table 3.7 – Summary of lifetime parameters obtained for 2AP-containing dinucleotides in water. Lifetime parameters are estimated to have the following error: $\tau_1 \leq 12\%$, $\tau_2 \leq 5\%$, $\tau_3 \leq 2\%$, $\tau_4 \leq 2\%$, $A_1 \leq 10\%$, $A_2 \leq 4\%$, $A_3 \leq 2\%$, and $A_4 \leq 2\%$.

All of the dinucleotides display the typical spread of lifetimes associated with 2AP incorporated into a nucleic acid construct (τ_1 < 0.1 ns, τ_2 ~0.5 ns, τ_3 ~2 ns, and τ_4 ~9 ns). Although lifetimes are generally similar between the different dinucleotides there are some features that should be emphasised. For the τ_1 lifetime, 2AP-G and 2AP-A exhibit the shortest and longest values, respectively, of the samples investigated. This follows from the expected charge transfer efficiency of the bases and is consistent with the results of Fiebig, Wan, and Zewail, which were at much higher time-resolution.[6] The longest lifetime, τ_4, shows the smallest percentage change between the different dinucleotides. As this lifetime is attributed to an open conformation, where bases are unstacked and separated from each other, it is logical that there is little dependence on the identity of the neighbouring base. The slight reduction in lifetime from ~11.5 ns for free 2AP to ~9 ns in the dinucleotides is comparable to the change observed above (Figure 3.15) due to the presence of the sugar group.

It is difficult to fully account for the observed A-factor behaviour but, again, there are particular points that should be highlighted. The magnitude of the A_1 component is largest in 2AP-G and smallest in 2AP-A. This could reflect upon the base-stacking efficiency of the bases but it is much more likely to be another manifestation of the relative charge transfer ability of the bases. This can be explained by considering the fact that stacking does not need to be as efficient in 2AP-G for charge transfer to occur and so there is a greater probability for this dinucleotide to be in a conformation susceptible to this relaxation mechanism at any given time.

A_2 is considerably higher in 2AP-A (and slightly higher in 2AP-I) than the other dinucleotides, which show almost equivalent fractional contributions. A possible explanation for this observation is discussed later.

In all dinucleotides there is only a very small proportion (< 10%) of the emitting populating which exhibits a lifetime similar to that of free 2AP. This implies that there are relatively few instances where the 2AP moiety does not interact with the rest of the dinucleotide structure on the timescale of this lifetime. 2AP-I has the largest A_4 component, which correlates to the fact its associated decay shows the slowest attenuation overall. This would imply that the relative population of the open dinucleotide conformation is greatest in 2AP-I. The same reasoning would suggest that 2AP-T has the smallest relative population of dinucleotides in an open conformation; however, these observations should be treated with caution. A-factors do not necessarily provide a fair representation of the conformational distribution in a system because the proportion of each component can depend (sometimes fully) on factors other than conformation (as highlighted with charge transfer in 2AP-G). The interdependence of A-factors also means that errors are propagated through all values and so care should be taken when evaluating the importance of small differences in small values.

A more detailed consideration of the fit parameters highlights the fact that A_1 and A_2 components combined account for around 75% of the emitting population for 2AP-A and 2AP-G; this fraction is significantly lower for 2AP-T (64%) and 2AP-C (53%). This implies that the most efficient quenching mechanisms are more readily accessible with a proximal purine base than a pyrimidine base. A possible explanation for this is more efficient stacking behaviour between bases of the same type. Chapter 5 explores this possibility and the results suggest that there is slightly different conformational behaviour in stacking between 2AP and the two base types; however, it is worth noting that the short lifetime components only account for 55% of the emitting population of 2AP-I, despite hypoxanthine (inosine) being purine based. Therefore factors other than conformation must also contribute to differences observed in fractional contribution of the shortest lifetime components. Charge transfer efficiency could explain this discrepancy for 2AP-G but it cannot account for the fact that the two shortest lifetime components also dominate the 2AP-A fluorescence decay. Computational calculations of 2AP-containing dimers by Jean and Hall offer further insight into the differences observed for purine and pyrimidine neighbours.[36] Their results showed that mixing of molecular orbitals in the ground state caused static quenching when 2AP was stacked

with a purine base. On the other hand, formation of a low-lying dark excited state enabled dynamic quenching when 2AP was stacked with a pyrimidine base. The availability of different quenching mechanisms was predicted to be cause measurable differences in fluorescence lifetimes and quantum yields.

Overall the parameters obtained are generally consistent with previous studies of 2AP-containing dinucleotides by Somsen *et al.*[12] and Wu[15] (a direct comparison of parameters is given in Appendix 3, §3.2); however, the sub-unity χ_R^2 values suggest that the curves, with the exception of 2AP-G, may have been fitted using too many parameters. Indeed, as previously mentioned, Wu found that 2AP-A and 2AP-I could be adequately described by a three-exponential model;[15] this could explain the previous observation of the anomalously high A_2 component of these dinucleotides. Specifically, it would appear that the A_1 and A_2 terms can be described by a single component. This does not necessarily mean that one of the components is completely absent; it could simply mean that, within the experimental resolution, it is not possible to distinguish both components. The results of Fiebig, Wan, and Zewail corroborate that indistinguishability is a reasonable proposal in this case.[6] The rapid decay component of 2AP-A was observed to be 350 ps; this is significantly longer than the other dinucleotides (the closest was 2AP-I at 120 ps) and is approaching the typical lifetime of the second component in these systems (~500 ps). It therefore seems plausible that the first and second components could be adequately described by a single decay (assuming the lifetime of the second component remains around 500 ps in 2AP-A). The rapid decay component of 2AP-I was intermediate between 2AP-A and the other dinucleotides and so this explains why the A_2 magnitude was only slightly perturbed in this dinucleotide.

At this point it is worth clarifying why a four-exponential model was preferred over a three-exponential model for the dinucleotide decays measured in this study. As an illustrative example, the decay recorded for the 2AP-T dinucleotide in pure water was fitted with a three-exponential function (Figure 3.19) and a four-exponential function (Figure 3.20); fit parameters are given in Table 3.8

Decay	Lifetimes /ns				A-Factors				χ_R^2	$\langle\tau\rangle$ /ns
	τ_1	τ_2	τ_3	τ_4	A_1	A_2	A_3	A_4		
3 Exp	-	0.17	2.27	7.90	-	0.35	0.61	0.04	1.135	1.745
4 Exp	0.05	0.66	2.45	9.36	0.55	0.09	0.34	0.01	0.987	1.070

Table 3.8 – Three-exponential (3 Exp) and four-exponential (4 Exp) fit parameters obtained for the fluorescence decay of 2AP-T in pure water at 22°C. Sample concentration was ~10-20 µM. The 3 Exp model was rejected as a possible model of the data due to a high χ_R^2 value and the non-randomness of the fit residuals (see Figure 3.19).

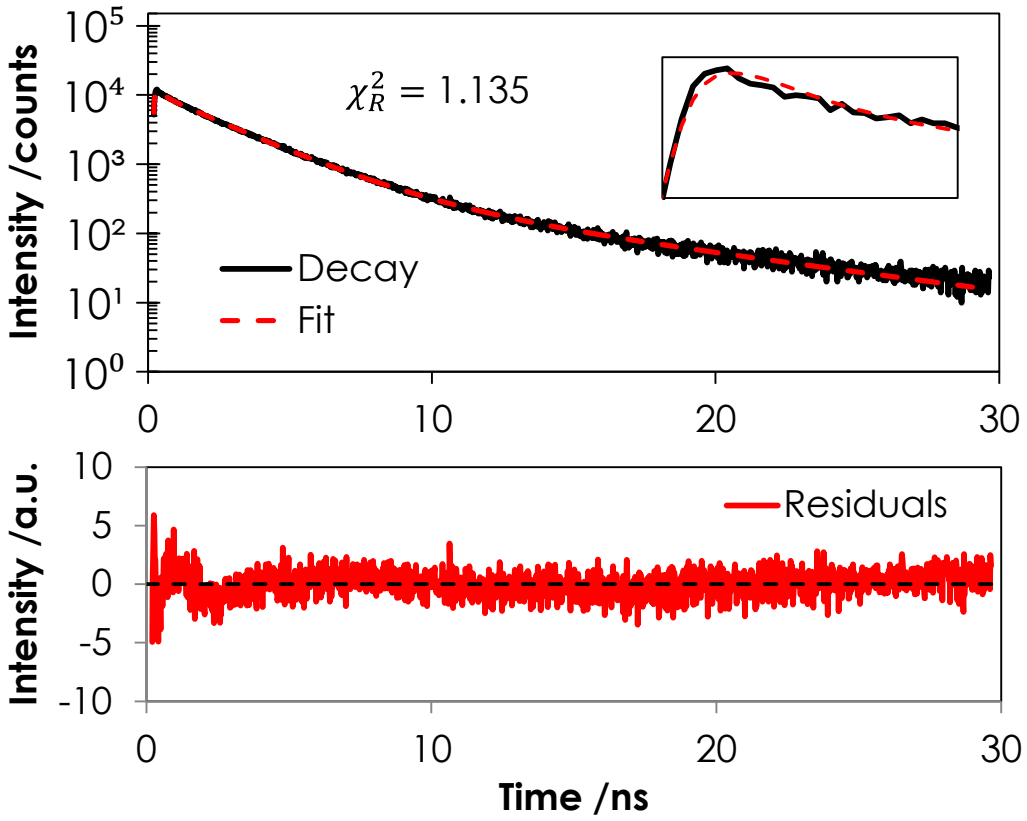

Figure 3.19 – (Top) Three-exponential fit of the fluorescence decay of 2AP-T in pure water at 22°C. Sample concentration was ~10-20 µM. The inset shows early times (0.2 to 0.7 ns) on a linear intensity scale. (Bottom) Weighted residuals of the fit compared to the decay.

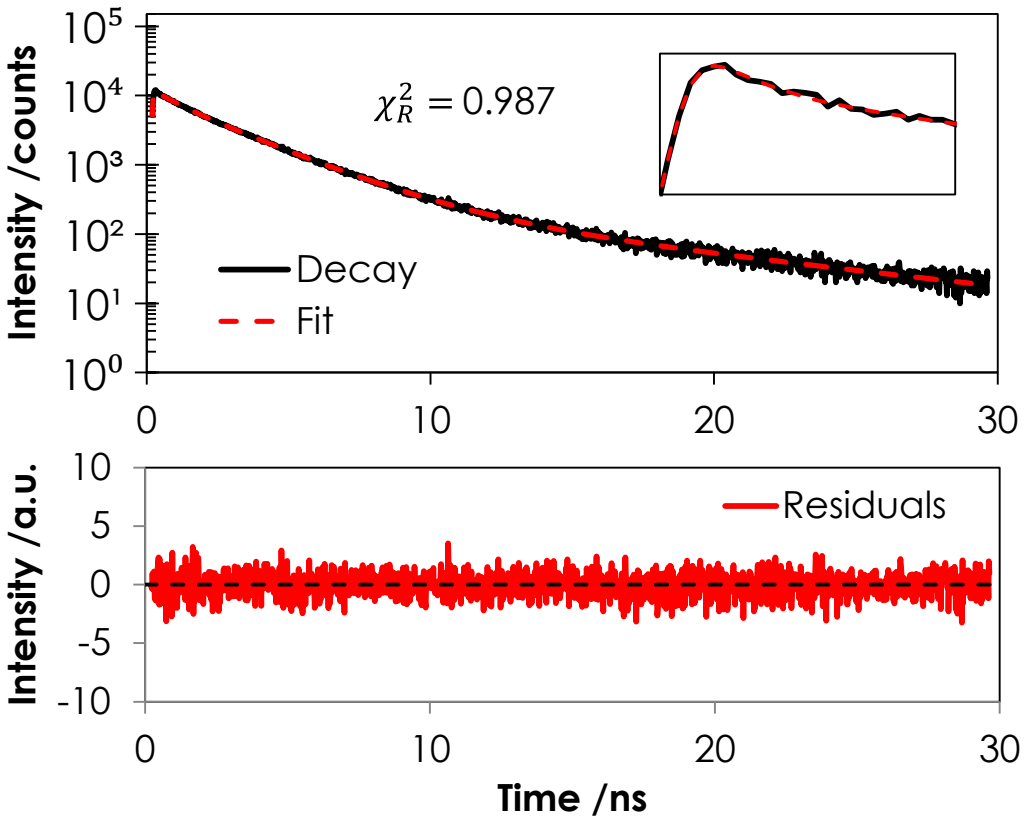

Figure 3.20 – (Top) Four-exponential fit of the fluorescence decay of 2AP-T in pure water at 22°C. Sample concentration was ~10-20 μM. The inset shows early times (0.2 to 0.7 ns) on a linear intensity scale. (Bottom) Weighted residuals of the fit compared to the decay.

Although the χ_R^2 value is fairly close to 1.0, the non-random distribution of the residuals indicates that a three-exponential fit (Figure 3.19) does not adequately describe the data. The inset, which shows early times, highlights the discrepancy between the measured decay and the fit. In contrast, the near-unity χ_R^2 value and the random distribution of the residuals indicate that a four-exponential fit (Figure 3.20) adequately describes the data. The superiority of the four-exponential model over the three-exponential model at describing the measured decay was typical for all dinucleotides (including those in alcohol-water mixtures, which are presented below).

The consistent use of a four-exponential model for all dinucleotide samples allows a much more intuitive comparison to be made between them; however, the potential for anomalous effects caused by over-fitting should be taken into consideration during further discussion. A more rigorous analysis of the fluorescence decay of 2AP-containing dinucleotides is presented in the following chapter.

3.3.4.2 2AP-N Dinucleotides in Alcohol-Water Mixtures

The lifetime parameters obtained for the 2AP-N dinucleotides in various alcohol-water mixtures are shown below. Connecting lines are physically meaningless but are presented to guide to eye. All fit parameters can be found in Appendix 3, §3.2.

Lifetimes and A-factors are plotted against the mole fraction of alcohol. This contrasts the study by Nakabayashi, Islam, and Ohta where the dielectric coefficient of the solvent was used as the abscissa;[16] however, discontinuities were found between solvent types in this study and so it did not seem appropriate (or meaningful) to follow their methodology here. Alcohol mole fraction was chosen as a simple measure of the mixture composition as no other parameter was found that could fully account for the observed behaviour.

Lifetime values are plotting on a logarithmic scale to allow trends to be observed at all magnitudes. Related parameters (A_i and τ_i) have been matched in colour and marker type. For clarity error bars are not shown but lifetime parameters are estimated to have the following error: $\tau_1 \leq 20\%$, $\tau_2 \leq 12\%$, $\tau_3 \leq 10\%$, $\tau_4 \leq 2\%$, $A_1 \leq 25\%$, $A_2 \leq 12\%$, $A_3 \leq 15\%$, and $A_4 \leq 15\%$. An error of $\leq 5\%$ is estimated for alcohol content. As a reminder, all dinucleotide decay curves were analysed using a four-exponential model. Due to its unusual behaviour, 2AP-A will be presented last and only discussed briefly here. A more detailed analysis can be found in Chapter 4.

Figure 3.21 shows lifetime parameters obtained for 2AP-C in glycerol-water (filled markers) and ethanol-water (empty markers) mixtures.

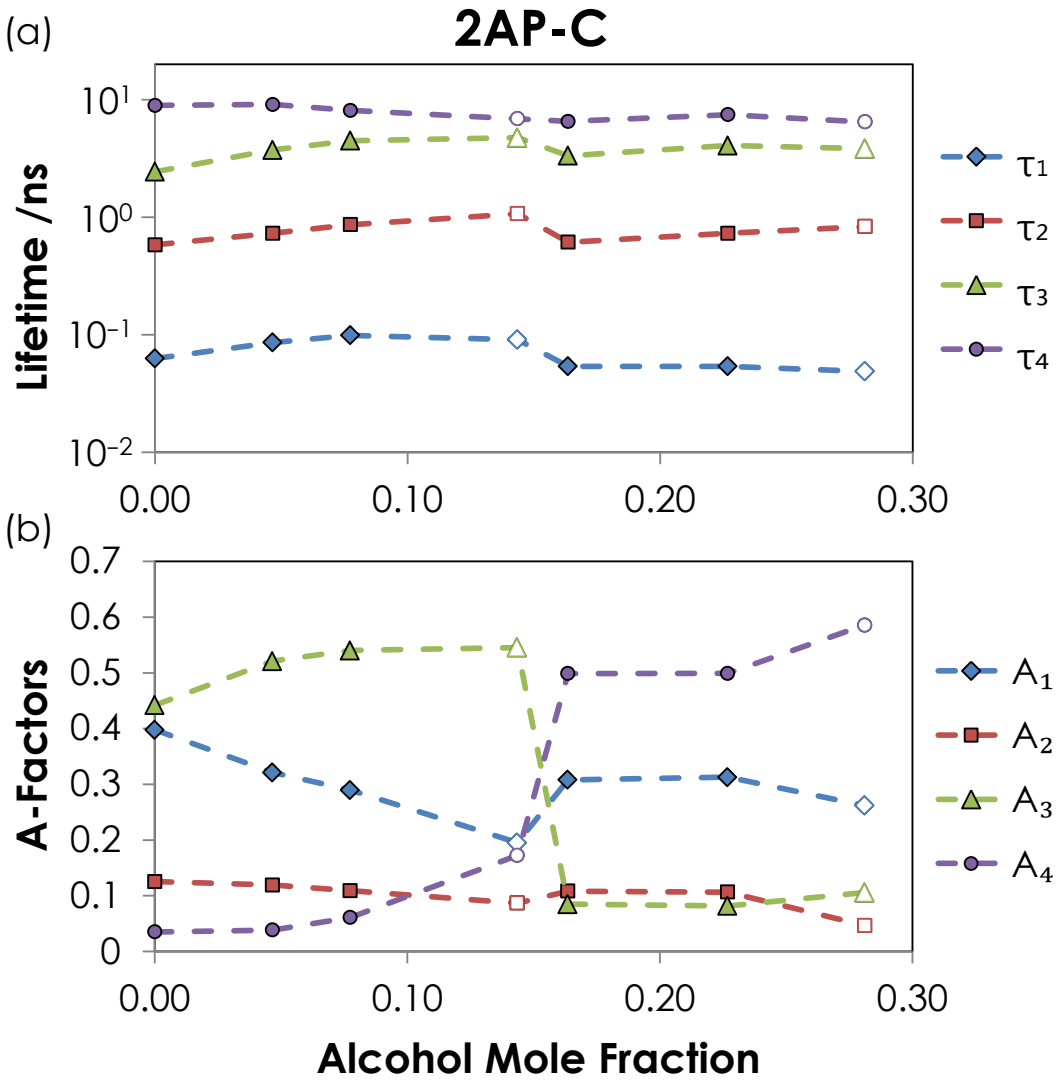

Figure 3.21 – Summary of (a) lifetimes and (b) A-factors obtained for 2AP-C in various alcohol-water mixtures. The parameters are plotted against mole fraction of alcohol. In all cases sample concentration was ~10-20 µM. Filled markers represent glycerol-water mixtures (including water only) while empty markers represent ethanol-water mixtures.

It is evident that the fluorescence character of the 2AP-C dinucleotide is sensitive to the solvent mixture. Lifetime values show only slight variation across all of the samples (though the changes were sufficient to prohibit global analysis). The longest lifetime generally decreases with increasing alcohol content (from ~9 ns in pure water to ~6.5-7.0 ns at highest alcohol content), which is consistent with the observations made for free 2AP. The τ_3 and τ_4 lifetimes are quite similar at higher alcohol content, perhaps suggesting that they could be modelled by a single lifetime; however, fitting with only three lifetimes typically resulted in a poor distribution of the residuals at early times for this dinucleotide.

The changes to A-factors are much more complicated. A_1 shows a slight decrease with alcohol content while A_2 remains relatively constant across all samples. Interestingly, the reduction in A_1, particularly at low alcohol content, appears to be mainly compensated by an increase in A_3 only (rather than an even increase in A_2, A_3, and A_4). Components A_3 and A_4 show correlated behaviour as changes in one are mirrored by the other; as one decreases the other increases (and *vice versa*).

Curiously, the trends in behaviour seem to be independent of alcohol type when plotting against mole fraction. This is somewhat unexpected given the significantly different properties of glycerol and ethanol but, as suggested previously, could be a result of the fact that the alcohol content is quite low in all samples and the local environment might always be dominated by water. It is also possible that this observation is purely coincidental and a result of having only two data points for the ethanol series. Therefore, for the sake of simplicity, the remaining plots of lifetime parameters will only include the glycerol-water mixtures.

Figure 3.22, Figure 3.23, and Figure 3.24 show the lifetime parameters obtained for 2AP-G, 2AP-T, and 2AP-I, respectively.

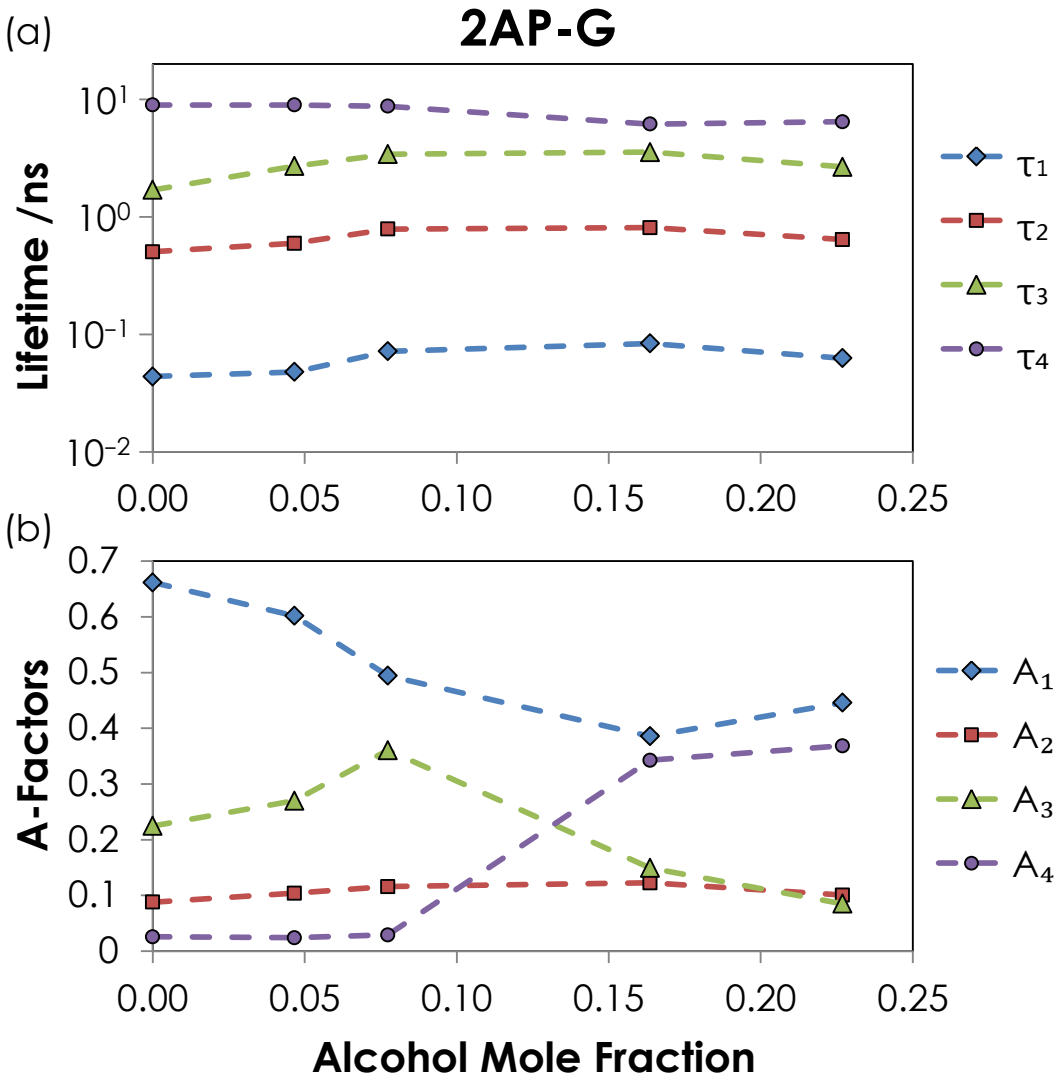

Figure 3.22 – Summary of (a) lifetimes and (b) A-factors parameters obtained for 2AP-G in various glycerol-water mixtures. Sample concentration was ~10-20 µM in all cases.

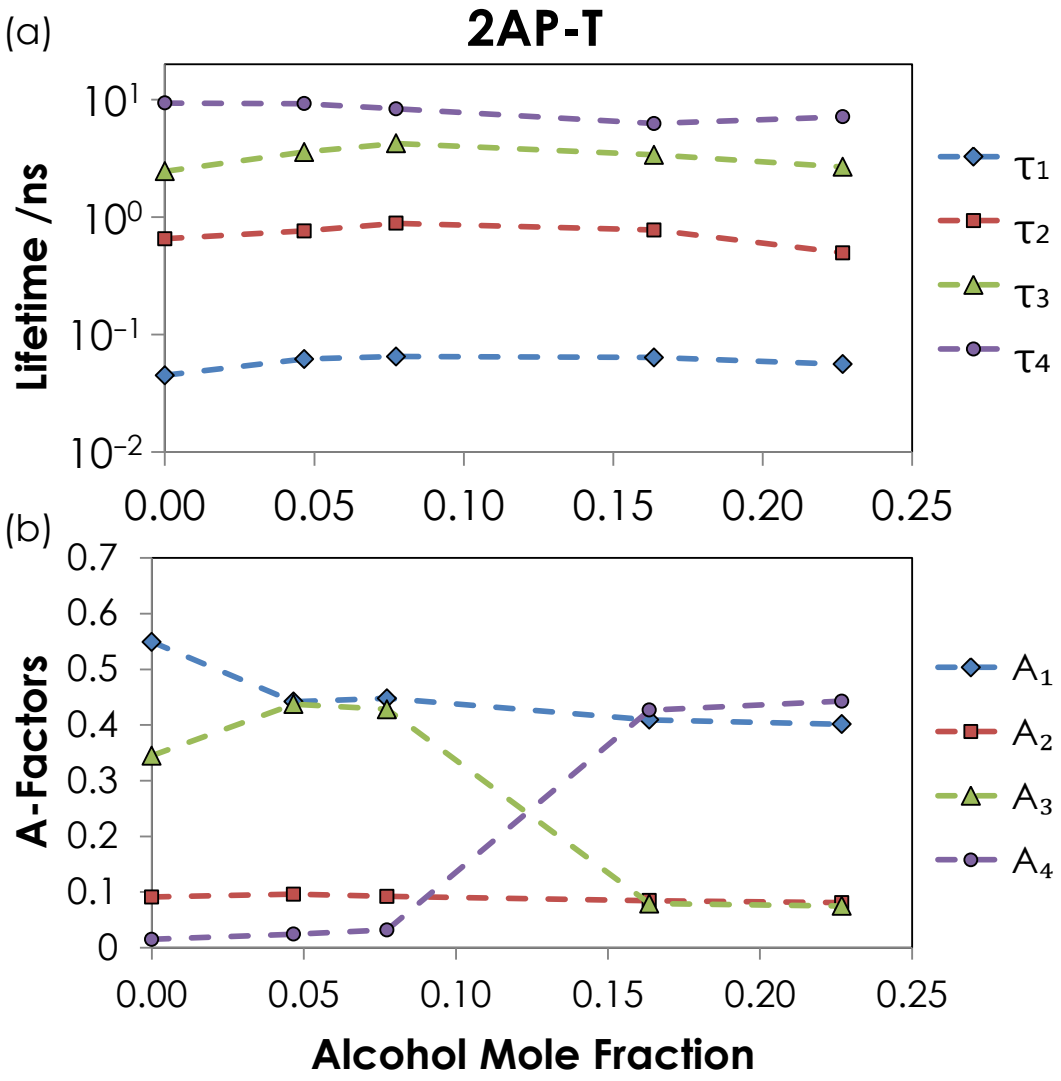

Figure 3.23 – Summary of (a) lifetimes and (b) A-factors obtained for 2AP-T in various glycerol-water mixtures. Sample concentration was ~10-20 μM in all cases.

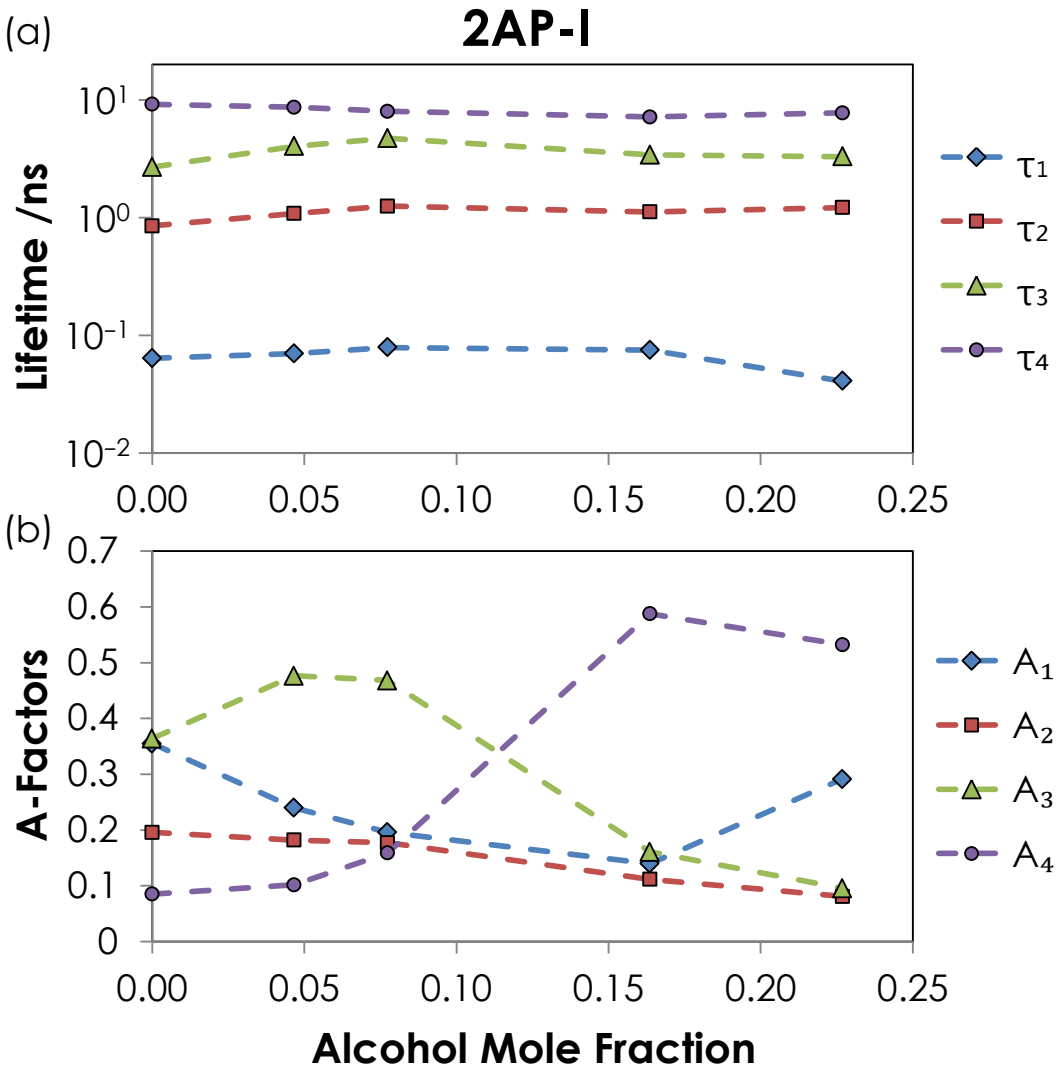

Figure 3.24 – Summary of (a) lifetimes and (b) A-factors obtained for 2AP-I in various glycerol-water mixtures. Sample concentration was ~10-20 μM in all cases.

The general changes observed in 2AP-C are also seen in these dinucleotides: lifetime values remain fairly constant (although there is typically a reduction in τ_4 of around 20-30%); A_1 shows a gradual decrease that is compensated by A_3 only; A_3 and A_4 mirror each other; and A_2 stays at approximately the same value across all samples.

The main differences observed in the behaviour of the dinucleotides can be attributed to the magnitude of the short lifetime component (A_1). This generally follows the trend 2AP-G > 2AP-T > 2AP-C > 2AP-I, which is consistent with the predicted charge transfer efficiency of the base neighbouring 2AP. The decrease in A_1 caused by addition of alcohol can be explained by the fact that base-stacking will become less favourable and so bases are more likely to be separated from each other beyond the threshold for rapid charge transfer to occur.

As previously mentioned, it is possible that 2AP-I has been over-fitted for the resolution of the experimental method and this is possibly the cause of the relatively high magnitude of A_2 in this dinucleotide compared to the others. The mixed nature of the two short lifetime components might explain why there is a slight reduction in A_2 as glycerol content increases, an effect not observed in the other dinucleotides. There also appears to be an anomalously high A_1 component at the highest alcohol content shown (though it should be noted that the E50 ethanol-water mixture also showed this behaviour).

Figure 3.25 shows lifetime parameters obtained for 2AP-A; it is immediately evident that the behaviour of this dinucleotide is quite distinct from those previously discussed.

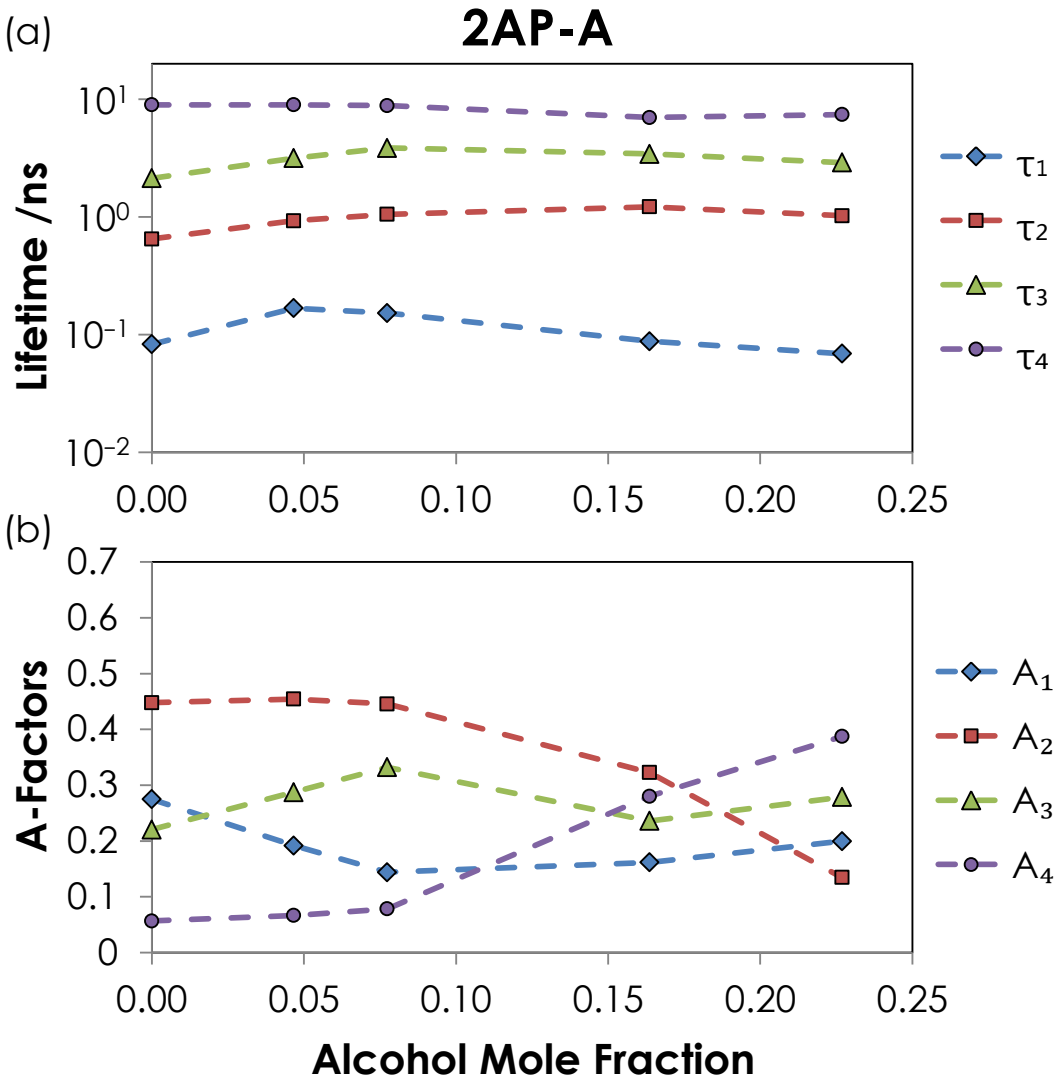

Figure 3.25 – Summary of (a) lifetimes and (b) A-factors parameters obtained for 2AP-A in various glycerol-water mixtures. Sample concentration was ~10-20 µM in all cases.

2AP-A lifetime values are again fairly constant for all of the samples studied; however, the corresponding A-factors exhibit unique behaviour. In contrast to all other dinucleotides, the A_2 component shows significant dependence on solvent composition. The A_2 component is large in pure water, consistent with previous observations of this dinucleotide,[12] but as alcohol content increases there is a gradual decrease in its occupancy. A_1 and A_3 are fairly constant across all samples but, similar to the other dinucleotides, appear interrelated (particularly at low alcohol content); as A_1 decreases there is an increase in A_3 (A_1+A_3 is approximately constant). The relative population of A_4 follows the same trend as for other dinucleotides (low at low alcohol content and high at high alcohol content) but the typical interrelationship with A_3 is not evident.

Given the discrepancy that has been observed in previous studies, it is possible that the unusual behaviour of 2AP-A arises because the assumed decay model is not appropriate for this dinucleotide. The 2AP-I dinucleotide parameters were also likely affected by the same issue but not as significantly as 2AP-A. Some more insight into this issue will be presented in Chapter 4.

Figure 3.26 shows the amplitude-weighted lifetime, $\langle\tau\rangle$, obtained for all of the alcohol-water mixture samples.

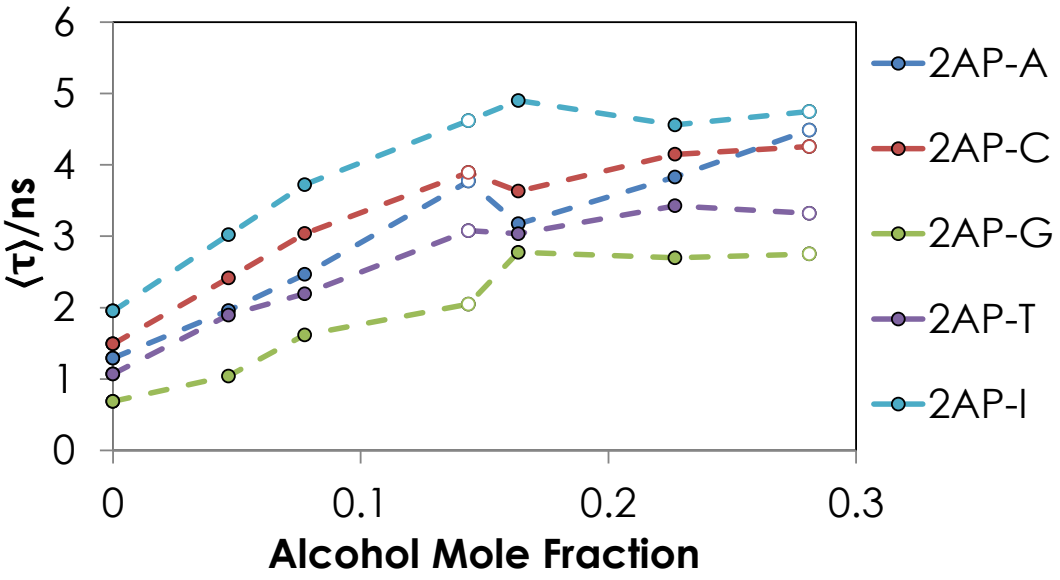

Figure 3.26 – Amplitude-weighted lifetime, $\langle\tau\rangle$, of 2AP-containing dinucleotides in various glycerol-water (filled markers) and ethanol-water (open markers) mixtures. Sample concentration was ~10-20 µM in all cases.

Despite the lifetime of free 2AP decreasing with increasing alcohol content (as observed in Table 3.6 and Figure 3.17) there is a general increase in the

amplitude-weighted lifetime of the 2AP-containing dinucleotide. This apparent contradiction can be resolved by considering the change in the population of the lifetime components when changing the solvent; specifically, the shift in population to longer-lived components far outweighs the decrease in lifetimes and causes a significant increase in amplitude-weighted lifetime. There is potentially a plateau in amplitude-weighted lifetime at higher alcohol content. Although somewhat speculative, given the limited number of data points, this proposal could be physically realistic if there is an alcohol content threshold, beyond which there is no longer any change in the relative proportions of dinucleotides in stacked and open conformations.

3.3.5 Implications of the Solvent-Dependence of Dinucleotide Decay Parameters

The gradual decrease in the longest dinucleotide lifetime with increasing alcohol content (up to around 20-30% at the highest alcohol content used) is consistent with the interpretation that this component is due to 2AP which is in an environment similar to that experienced by free 2AP. It seems plausible that changes in alcohol content would also have an effect on the other lifetimes too; however, the interdependence of the fitting parameters makes it difficult to see clear trends in behaviour.

Conformational dynamics has been found to play a significant role in determining the observed fluorescence behaviour. In this interpretation, the switch in relative magnitude of the A_3 and A_4 components can be attributed to a shift in conformational equilibrium towards opens states as the polarity of the solvent mixture decreases.

The insensitivity of A_2 to solvent conditions is very interesting and could have important consequences for the interpretation of the fluorescence lifetime data from 2AP in nucleic acids. There is still significant uncertainly regarding the possible cause of the intermediate lifetimes in these systems because, unlike the stacked and open conformations, there are no obvious structures that can be intuitively held accountable. This had led to speculations about whether or not they really are the result of distinct, independent conformations.

One approach has been to model the system as having well-defined stacked and open structures while invoking a broad distribution of conformations to account for the intermediate lifetime.[37] This seems physically sensible as one can envision a range of conformations with varying degrees of intermolecular interactions present; however, while being an elegant solution, this distributive approach is at odds with the observations made here. If the intermediate components were really a simplistic

representation of a distribution of lifetimes then there should be correlation between the associated A-factors (A_2 and A_3), which is not observed.

Another approach to explaining the presence of intermediate lifetimes is to include interconversion between distinct excited states. Transition to a dark state[4,5] and conformational gating[12,13] have been proposed as possible mechanisms for this behaviour. Interestingly, the reduction in A_1 appears to be compensated by an increase in A_3 only (rather than an even increase in A_2, A_3, and A_4). This is consistent with the three-state model proposed by Somsen *et al.*[12] In their model the excited states associated with components A_1 and A_3 were predicted to originate from the same ground state. These components were also found to be more temperature dependent than the other two; this substantiates the idea that the A_1 and A_3 components have conformational dependence. Speculatively, the conformation related to A_3 could be a state where the twist between bases is hyper-rotated (Chapter 5). In this conformation, the presence of an H-Bond with the sugar group of the 3'-base relative to 2AP could be factor involved in a gating mechanism. A_3 might also represent a conformation similar to the base pair opening state that is required for imino proton exchange in a duplex; for exchange to occur the base must rotate by around 30° to allow solvent access. A distorted stacked state has also been suggested as a possible preliminary stage to base-flipping.[38] It seems plausible that a perturbed stacked conformation might require rearrangement to permit charge transfer, which is known to be sensitive to structural dynamics.[9,39]

Another potential cause of the complexity of the fluorescence decay is the availability of multiple relaxation pathways in the excited state. Computational studies by the Matsika group[40,41] have shown that the pathway favoured is dependent on the initial conformation of the excited state as well as the identity of the base neighbouring 2AP; this is discussed in more detail in Chapter 5.

While the results of the present study have provided some useful insights into the relationship between the observed lifetimes component, there is clearly still a need for further investigation to establish the precise causes of the complicated fluorescence behaviour of 2AP in nucleic acids.

3.4 Conclusions

Steady-state and time-resolved fluorescence measurements have been performed on 2AP and 2AP-containing dinucleotides in various alcohol-water mixtures to gain insight into the effect of solvation on the fluorescence behaviour of these systems. A brief summary of the findings is presented below.

3.4.1 2AP Fluorescence in Solution

Steady-state measurements have shown that spectral shifts result as a consequence of changing the solvent environment of 2AP. These findings have been rationalised by considering solvent polarity and solvent reorganisation effects. In addition to affecting steady-state spectra, it is clear that solvent polarity also has a significant effect on the fluorescence lifetimes of 2AP in solution; however, the results from alcohol-water mixtures suggest that there must be other factors that contribute to the precise photophysical behaviour observed. Further investigation is required to elucidate the full influence of solvation on the fluorescence of 2AP. Solvation effects stand as the basis of the current understanding of the fluorescence behaviour of 2AP in nucleic acid systems and so it is vitally important that a more complete description is achieved.

3.4.2 2AP-Containing Dinucleotides in Water

Absorption spectra for 2AP-N dinucleotides were found to be consistent with the combined absorption profile of component bases. Selective excitation of 2AP can be achieved due to the negligible absorption of the natural bases at wavelengths longer than about 300 nm.

The fluorescence spectra for all 2AP-N dinucleotides were, in general, very similar to each other. The slight shifts in spectral peaks compared to free 2AP could be explained by considering the proximal base as a barrier against solvent interactions. The only significant difference was observed for the excitation spectrum of 2AP-A, which showed unusually high intensity at around 260 nm. This effect was attributed to the higher energy transfer efficiency of adenine to 2AP compared to the other bases.

Lifetime parameters obtained from time-resolved measurements displayed the typical characteristics expected for 2AP incorporated into a nucleic acid construct. Lifetime values were found to be similar between all dinucleotides but the contribution of the associated states to the emission decay showed greater variation. The identity of the neighbouring base clearly influenced the observed photophysical behaviour. The

most rapid decay was found to be shortest and the most dominant for 2AP-G while being longest and least dominant for 2AP-A; the overall trend for the contribution of the short component term was G > T > C > I > A (order reversed for magnitude of the associated lifetime). These observations are consistent with the expected propensity for charge transfer to occur in these dinucleotides.

The lifetime parameters obtained were generally consistent with those obtained in earlier studies of 2AP-containing dinucleotides; however, Wu had found that a three-exponential model had been adequate for 2AP-A and 2AP-I.[15] Use of an inappropriate decay model could possibly account for the unusual distribution of A-factors observed for 2AP-A (and to a lesser extent 2AP-I).

3.4.3 2AP-Containing Dinucleotides in Alcohol-Water Mixtures

Spectral shifts in steady-state measurement of the dinucleotides were found to follow the same trends observed for free 2AP. Interestingly, the energy transfer efficiency of 2AP-A was observed to decrease with the inclusion of alcohol; this was rationalised by considering the shift in conformational equilibrium towards unstacked states as solvent polarity decreased.

The lifetime parameters obtained from time-resolved fluorescence measurements generally showed the same behaviour for all dinucleotide. The longest lifetime became shorter as alcohol content increased, consistent with the behaviour of free 2AP, but it was difficult to extract any meaningful trends in the other components due to their interdependence. The overall behaviour of the A-factors suggested that there was an interrelationship between A_1, A_3, and A_4 but not A_2. The independence of A_2 might indicate that this component corresponds to a distinct relaxation mechanism from the other components, which appear to be mainly dominated by charge transfer effects. This interpretation is corroborated to some extent by the fact that the value of A_2 was generally independent of identity of the neighbouring base. It would appear that the initial addition of alcohol mainly impacted the conformational stability of the fully stacked dinucleotide form, thought to undergo efficient quenching by rapid charge transfer. Consequently, there was an increase in the relative population of the longer-lived A_3 state, which could potentially be associated with a partially stacked conformation that requires structural rearrangement to relax *via* charge transfer. Further addition of alcohol shifted the conformational equilibrium from stacked conformations towards open conformations. This shift resulted in a dramatic decrease in the population of A_3 which was compensated by an increase in the A_4 component,

which is associated with free-like 2AP. Curiously, A_1 remained fairly constant during this transition and, at the highest alcohol content investigated, the dinucleotides were most likely to be efficiently stacked or fully open; in other words, they favoured the extremes of possible conformations. The qualitative differences in lifetime parameters could mostly be explained by the same arguments made for the pure water samples, which mainly concerned charge transfer character. 2AP-A was found to display unusual trends in lifetime parameters but this could perhaps be attributed to the fitting procedure used. 2AP-I also showed slight signs of atypical behaviour but the effect was much less significant than observed for 2AP-A; nevertheless, this highlights the need for care when analysing complex systems.

3.4.4 Outlook

By altering the solvent environment of 2AP-containing dinucleotides it has been possible to change the distribution of conformational states present. The findings corroborate the current interpretation that structural dynamics play an important role in the complicated fluorescence behaviour observed in these systems.

It is again worth highlighting that the limited time available for time-resolved experiments meant that the analysis presented is typically based on just a single decay. In addition, decays were only recorded to a fairly low number of counts in the peak channel (10,000 counts); this may not have been sufficient for the decays with multiple lifetime components.[42,43] The precise value of the lifetime parameters presented should therefore be treated with caution, particularly for the more complicated dinucleotide decays. Nevertheless, some confidence in the results can be gained from the fact that the lifetime parameters obtained were consistent with previous studies. Furthermore, recording decays for a whole series of samples meant that the reliability of each result could, to some extent, be verified against a similar sample. The investigation was also more concerned with gaining insight into the qualitative trends in behaviour of 2AP photophysics in alcohol-water mixtures rather than obtaining precise lifetime values. As such, any deficiency in the protocol used for time-resolved measurements may have had limited influence on the conclusions made. This being said, it would likely be of great benefit to perform a much more rigorous study of the 2AP systems investigated in this chapter to ensure the trends observed are really trustworthy.

An obvious extension of the present study would be to investigate more solvents and cover a greater range of mixture composition. It would be particularly interestingly to examine solvents that had varying degrees of H-bonding ability.

The crossover for A_3 and A_4 components always occurred between the same two glycerol-water mixture compositions (30% and 50% (w/w) glycerol) and so it would be interesting to investigate a series of compositions between these points to determine a more accurate threshold composition for each dinucleotide; differences in threshold compositions might provide insight into the relative strength of the base-stacking present in each dinucleotide. It is easy to see how computational studies might be complementary here; molecular dynamics simulations might be able to elucidate the local solvation environment in these systems and provide some insight into why there appears to be a threshold composition in the first place.

It would also be interesting to investigate the effect of changing the solvent composition on duplex systems. While more complicated, it would be useful to see if similar effects were observed on much larger systems.

These examples highlight the fact that there are clear opportunities for improving the understanding of solvation effects on 2AP fluorescence (both in isolation and in nucleic acid environments). As 2AP is widely used to interrogate complex processes that are crucial to biological function, it is imperative that an accurate description of these effects be obtained.

3.5 References

1. C. Reichardt, C. Wen, R. A. Vogt, and C. E. Crespo-Hernández, *Photochem. Photobiol. Sci.*, 2013, **12**, 1341–50.

2. R. K. Neely, S. W. Magennis, D. T. F. Dryden, and A. C. Jones, *J. Phys. Chem. B*, 2004, **108**, 17606–10.

3. R. Berera, R. van Grondelle, and J. T. M. Kennis, *Photosynth. Res.*, 2009, **101**, 105–18.

4. O. F. A. Larsen, I. H. M. van Stokkum, M.-L. Groot, J. T. M. Kennis, R. van Grondelle, and H. van Amerongen, *Chem. Phys. Lett.*, 2003, **371**, 157–63.

5. O. F. A. Larsen, I. H. M. van Stokkum, F. L. de Weerd, M. Vengris, C. T. Aravindakumar, R. van Grondelle, N. E. Geacintov, and H. van Amerongen, *Phys. Chem. Chem. Phys.*, 2004, **6**, 154–60.

6. T. Fiebig, C. Wan, and A. H. Zewail, *Chemphyschem*, 2002, **3**, 781–8.

7. C. A. M. Seidel, A. Schulz, and M. H. M. Sauer, *J. Phys. Chem.*, 1996, **100**, 5541–53.

8. S. Steenken and S. V. Jovanovic, *J. Am. Chem. Soc.*, 1997, **7863**, 617–8.

9. S. O. Kelley and J. K. Barton, *Science*, 1999, **283**, 375–81.

10. C. E. Crespo-Hernandez, D. M. Close, L. Gorb, and J. Leszczynski, *J. Phys. Chem. B*, 2007, **111**, 5386–95.

11. M. Narayanan, G. Kodali, Y. Xing, and R. J. Stanley, *J. Phys. Chem. B*, 2010, **114**, 10573–80.

12. O. J. G. Somsen, L. B. Keukens, M. N. de Keijzer, A. van Hoek, and H. van Amerongen, *Chemphyschem*, 2005, **6**, 1622–7.

13. M. A. O'Neill, H.-C. Becker, C. Wan, J. K. Barton, and A. H. Zewail, *Angew. Chem. Int. Ed. Engl.*, 2003, **42**, 5896–900.

14. M. A. O'Neill and J. K. Barton, *J. Am. Chem. Soc.*, 2004, **126**, 11471–83.

15. X. Wu, PhD Thesis, University of Edinburgh, 2012.

16. T. Nakabayashi, Md. Serajul Islam, and N. Ohta, *J. Phys. Chem. B*, 2010, **114**, 15254–60.

17. L. Levine, J. A. Gordon, and W. P. Jencks, *Biochemistry*, 1963, **2**, 168–75.

18. G. Baldini, H. Fu-Hua, G. Varani, L. Cordone, S. L. Fornili, and G. Onori, *Nuovo Cim. D*, 1985, **6**, 618–30.

19. A. Albert and H. Taguchi, *J. Chem. Soc. Perkin Trans. 2*, 1973, 1101–3.

20. J. Smagowicz and K. L. Wierzchowski, *J. Lumin.*, 1974, **8**, 210–32.

21. L. M. Wilhelmsson, *Q. Rev. Biophys.*, 2010, **43**, 159–83.

22. M. J. Cavaluzzi and P. N. Borer, *Nucleic Acids Res.*, 2004, **32**, e13.

23. G. Åkerlöf, *J. Am. Chem. Soc.*, 1932, **54**, 4125–39.

24. K. Evans, D. Xu, Y. Kim, and T. M. Nordlund, *J. Fluoresc.*, 1992, **2**, 209–16.

25. C. Reichardt, *Chem. Rev.*, 1994, **94**, 2319–58.

26. MATLAB, *Version 8.2.0 (R2013b)*, The MathWorks, Inc., Natick, Massachusetts, United States of America, 2013.

27. Microsoft Excel (2010), *Version 14.0*, Microsoft Corporation, Microsoft Redmond Campus, Redmond, Washington, United States of America, 2010.

28. FAST, *Version 3.4.1*, Edinburgh Instruments Ltd., Kirkton Campus, Livingston, United Kingdom, See; http://www.edinburghphotonics.com/, 2013.

29. J. R. Lackowicz, *Principles of Fluorescence Spectroscopy*, Springer Science & Business Media, LLC, 3rd edn., 2006.

30. T. M. Nordlund, D. Xu, and K. O. Evans, *Biochemistry*, 1993, **32**, 12090–5.

31. D.-G. Xu and T. M. Nordlund, *Biophys. J.*, 2000, **78**, 1042–58.

32. R. K. Neely, PhD Thesis, University of Edinburgh, 2005.

33. O. K. Abou-Zied, *J. Photochem. Photobiol. A Chem.*, 2013, **261**, 1–6.

34. D. R. Lide, Ed., in *CRC Handbook of Chemistry and Physics*, CRC Press, Inc., Boca Raton, Florida, United States of America, 74th edn., 1993, p. 6: 198.

35. N.-S. Cheng, *Ind. Eng. Chem. Res.*, 2008, **47**, 3285–8.

36. J. M. Jean and K. B. Hall, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 37–41.

37. A. C. Fogarty, A. C. Jones, and P. J. Camp, *Phys. Chem. Chem. Phys.*, 2011, **13**, 3819–30.

38. D. M. Matje, H. Zhou, D. A. Smith, R. K. Neely, D. T. F. Dryden, A. C. Jones, F. W. Dahlquist, and N. O. Reich, *Biochemistry*, 2013, **52**, 1677–85.

39. M. A. O'Neill and J. K. Barton, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 16543–50.

40. J. Liang and S. Matsika, *J. Am. Chem. Soc.*, 2011, **133**, 6799–808.

41. J. Liang, Q. L. Nguyen, and S. Matsika, *Photochem. Photobiol. Sci.*, 2013, **12**, 1387–400.

42. D. R. James and W. R. Ware, *Chem. Phys. Lett.*, 1985, **120**, 455–9.

43. D. V. O'Connor and D. Phillips, *Time-Correlated Single Photon Counting*, Academic Press, 1984.

Chapter 4: A Comparison of Analysis Techniques for Time-Resolved Fluorescence

Without careful consideration of the nature of the problem, deconvolution as an information-improving device can easily become an exercise in self-delusion.

– Knight and Selinger, *Spectrochimica Acta,* 1971, **27A**, 1223

4.1 Introduction

This chapter outlines a brief evaluation of the benefits and limitations of a selection of different methods that have been used to analyse time-correlated single photon counting (TCSPC) data. As highlighted by the complexity of the fluorescence decay behaviour of the 2AP-containing dinucleotides investigated in the previous chapter, there is a need for accurate data analysis techniques that are able to provide a reliable description of the system under study. Although there will be a particular focus on the TCSPC method, the outcomes are relevant to all time-resolved fluorescence (TRF) techniques.

To begin with, the challenges involved in analysis of TCSPC data will be introduced. This will be followed by a summary of a few of the approaches that have been developed to overcome these difficulties. Later sections of this chapter will investigate how some of these methods perform with real and simulated data. This chapter will not rigorously cover the mathematical background which underpins the techniques used; a more detailed theoretical analysis can be found in the references provided throughout. It should also be noted that this study only represents a preliminary investigation into the comparison of the analysis techniques and will not review all possible methods in detail. Nevertheless, the results presented provide a basis for further studies and the analysis code that has been written should have future value (although some improvements will have to be made to address some of the deficiencies discussed below).

4.1.1 Deconvolution and an Ill-Posed Problem

The temporal resolution of a TCSPC experiment is limited by the instrument response function (IRF) of the system. The finite width of the excitation pulse, as well as the temporal broadening caused by the detector response, results in a perturbation to the measured decay that cannot be neglected. To be more specific, the measured decay intensity, $D(t)$, is a convolution of the IRF, $R(t)$, and the true fluorescence decay, $F(t)$;

$$D(t) = \int_0^t F(t')R(t-t')dt' \equiv [F * R](t). \qquad 4.1$$

In practice, because of the discrete nature of the channels used in the experiments, the convolution can be calculated as a summation;

$$D(t) = \sum_{t'=0}^{t} F(t')R(t - t'). \qquad 4.2$$

Regardless of the form of the convolution, the distortion of the true fluorescence decay by the IRF causes a serious problem for the analysis of TCSPC data. The difficulty lies with the fact that obtaining the true form of the fluorescence decay, $F(t)$, requires inversion of Equation 4.2, which is mathematically challenging because the problem is ill-posed.[1–3] Simply stated, this means that there are many solutions that exist that can adequately describe the observed decay behaviour. Under these conditions the presence of noise can have significant influence on the solution that is obtained. Ideally, it would be desirable to directly deconvolve the true fluorescence decay and the IRF from the measured decay. This would allow the true decay to be fitted in isolation and negate the perturbation caused by the IRF. Unfortunately, deconvolution strategies are generally found to be inappropriate for time-resolved fluorescence measurements.[4] For instance, techniques that utilise Fourier transforms suffer from discontinuities caused by truncation of the decay at long times (and also by the almost instantaneous excitation rise) in addition to the presence of noise.[5,6]

As well as the mathematical difficulty of inverting Equation 4.2, there is also the pragmatic problem of obtaining the IRF in the first place; irrespective of the analysis method used, an accurate measure of the IRF is essential if a reliable fit is to be achieved. Typically, the temporal profile of the IRF is determined by scattering the excitation pulse from a suspension of non-fluorescent particles, such as coffee creamer[7–10] or colloidal silica; however, this approach requires a detector that does not suffer from the *colour effect*, which is where the detector time response depends on the wavelength of the incident light.[11] Microchannel-plate photomultiplier tubes (MCP-PMTs), such as the one used for the experiments performed during this study, are essentially free of wavelength effects but avalanche photodiodes (APDs), which are commonly used for TCSPC because of high quantum yields, and conventional PMTs generally exhibit a wavelength dependence that cannot be neglected.[11,12] This issue can be addressed by using a reference fluorophore which has a very short, but known, lifetime and similar spectral properties as the sample.[11,13] In this case the IRF is implicit

in the reference lifetime and does not have to be directly measured. Of course, this method requires that the reference lifetime is well-established and does not change under the sample conditions used.

One of the drawbacks of measuring the IRF and the fluorescence decay separately is that there is potentially (and quite often) a small difference between the time origins of the experiments; this time difference will be referred to as the *q-shift* for the remainder of this chapter. To be able to accurately measure short lifetimes, which can be of similar magnitude as the width of each time channel, it is necessary to determine the q-shift to sub-channel precision.

An alternative approach to accommodating the IRF during fitting is to approximate it with an analytical function, such as a high-order polynomial.[14] This method can potentially give better time resolution and also benefits from fact that it is possible to perform analytical convolution, which may improve the reliability of the fit.

Irrespective of the approach used, the requirement for an accurate description of both the temporal profile of the IRF and its relative position in time compared to the measured fluorescence decay only adds to the difficulty of obtaining reliable fits of TCSPC data. Some methods that have been developed to overcome the challenges mentioned above will now be outlined.

4.1.2 Analysis Techniques for Time-Resolved Fluorescence

4.1.2.1 Iterative Reconvolution

Iterative (re)convolution[15] is the most commonly used analysis method for TCSPC measurements. The premise of this method is that a model function, typically of multi-exponential form, is convoluted with the measured IRF and fitted against the measured decay. Iteratively changing the variable parameters of the model function improves the fit. There are a number of commercial software packages that utilise this technique, including FAST[16] and FluorFit,[17] because it is a robust method that can quickly obtain solutions and, historically, it has been seen to be the most effective way to obtain reliable and accurate results.[4,5] During fitting the true fluorescence decay (which is typically normalised) is usually estimated by a lifetime distribution;

$$\frac{F(t)}{F_0} = F_N(t) = \int_0^\infty p(\tau) e^{-\frac{t}{\tau}} \mathrm{d}\tau, \qquad 4.3$$

where τ is the lifetime and $p(\tau)$ is the corresponding probability amplitude. For the common multi-exponential analysis $p(\tau)$ is represented by a weighted sum of *n* discrete delta functions;

$$p(\tau) = \sum_{i=1}^{n} a_i \delta(\tau - \tau_i), \tag{4.4}$$

which gives,

$$F_N(t) = \sum_{i=1}^{n} a_i e^{-\frac{t}{\tau_i}}, \tag{4.5}$$

where a_i is the contribution of the i^{th} lifetime under the normalisation condition;

$$\sum_{i=1}^{n} a_i = 1. \tag{4.6}$$

Usually the fitting process starts by using the simplest model and then progressively more complexity is added until an acceptable fit has been achieved. Once an adequate fit has been established there is generally no reason to believe that using a more complicated model would be beneficial or, in fact, appropriate.

The *goodness of fit* is generally determined using the (reduced) chi-squared statistic, χ_R^2;

$$\chi_R^2 = \frac{1}{\nu} \sum \frac{(O - E)^2}{\sigma^2} = \frac{1}{N_c - n_p - 1} \sum \frac{(D_N - D_{Calc})^2}{D_N}, \tag{4.7}$$

where $\nu = N_c - n_p - 1$ is the number of degrees of freedom in the fit (given by the number of channels, N_C, and the number of fitting parameters, n_p); O is the observed data (the measured decay, D_N); E is the estimate of the data (the calculated decay after convolution with the IRF, D_{Calc}) and σ^2 is the variance of the data (equal to D_N in Poisson statistics, which is typically assumed for fluorescence data). Roughly speaking the value of the χ_R^2 obtained can be interpreted in the following way; $\chi_R^2 \gg 1$ describes a poor fit; $\chi_R^2 > 1$ suggests that the model does not fully account for the observed behaviour;

$\chi_R^2 \cong 1$ indicates that, within the expected variance, the observed behaviour has been adequately matched by the model; finally, $\chi_R^2 < 1$ means that the data has been over-fitted, which can occur when there are too many variable parameters in the model, the error variance has been overestimated, or the fitting range is inappropriate.

One of the limitations of the iterative reconvolution method is that it requires a model to be assumed during the fitting process. Ideally the model used would be founded on some physical basis of the system but, unfortunately, there is often little or no knowledge that can aid in the construction of the proposed model. Without *a priori* knowledge, the model is usually accepted or rejected based on its statistical validity after fitting using the null hypothesis.[5] This approach initially assumes that all models are not valid; however, if the evaluated model meets some predetermined criteria (such as a χ_R^2 value less than some limit) it can no longer be discounted as a credible description of the system. It is important to realise that this does not necessarily mean that it is the correct model for the system; there may be a number of different models that satisfy the chosen criteria.

Global and Target Analysis

The reliability and robustness of a fit can be improved with the use of global analysis.[1,18,19] This approach requires simultaneous fitting of multiple decays that differ by some parameter; typically the excitation or, more commonly, the emission wavelength is altered. The fluorescence decay is then described by;

$$F_N(t,\lambda) = \sum_{i=1}^{n} a_i(\lambda) e^{-\frac{t}{\tau_i(\lambda)}} \cong \sum_{i=1}^{n} a_i(\lambda) e^{-\frac{t}{\tau_i}}, \qquad 4.8$$

where $a_i(\lambda)$ and $\tau_i(\lambda)$ are wavelength dependent amplitudes and lifetimes. Generally the lifetimes are assumed to be wavelength-independent (as rapid internal conversion means that, regardless of the excitation wavelength, fluorescence generally occurs from the same excited states) and so can be fixed when analysing multiple decays.

In some instances it is possible to predict (or estimate) the full underlying relaxation pathways available to the excited-state species. In such a case a kinetic scheme can be established that effectively stands as an interdependent system of linear differential equations.[20] Rather than calculate fluorescence decay lifetimes this strategy poses the problem as a series of excited-state reactions with the aim of determining rate constants; this approach is known as target analysis.[21,22] Solutions to these

problems can be found by numerical methods and are generally made more reliable by the use of global analysis. Transient absorption (although not a fluorescence technique) is very well-suited to this type of analysis because of the capacity for broadband detection, which generates a large number of decays at different wavelengths and, thus, maximises the benefit of global analysis. A significant advantage of target analysis is that it is able to handle complex relaxation dynamics. For instance, Robotham *et al.* employed global target analysis to investigate the relaxation dynamics in a porphyrin dyad and determined that relaxation to the S_1 excited-state could either occur directly from the S_2 excited-state or indirectly *via* a charge separated state;[23] this would be difficult to model using a standard exponential analysis approach. Global target analysis is clearly an extremely powerful technique; however, it requires considerable prior knowledge of the system to be able to construct an accurate kinetic scheme. Global and target analysis methods are not investigated in this thesis but for comprehensive reviews of these techniques see Ruckebusch *et al.*[24] and van Stokkum, Larsen, and van Grondelle.[20]

4.1.2.2 Probabilistic Analysis

The above methods can be considered deterministic techniques because the model which is eventually chosen is, generally, validated *a posteriori* by the quality of the fit. Unfortunately, this approach cannot definitively provide the correct model for the system and, if a number of possible solutions exist, it is not always clear which model is most physically appropriate; as was highlighted by the dinucleotide results in the previous chapter. It may therefore be more appropriate to use a probabilistic (or distributive) approach that makes no initial assumptions about the physical model of the system. It this case $p(\tau)$ in Equation 4.3 is not limited to any functional form. For this strategy it is more convenient to consider a *vectorised* formulation of the problem because of the large number of variables present. The true fluorescence decay can then be written in the following form;

$$\boldsymbol{F_N} = \boldsymbol{M}\boldsymbol{a}, \qquad 4.9$$

where $\boldsymbol{F_N}$ is the fluorescence decay binned in to m channels, $\boldsymbol{a}$ is the vector (length n) of A-factor amplitudes, $\boldsymbol{a} = [a_1, a_2, a_3, \dots, a_n]$, and $\boldsymbol{M}$, the so-called experimental matrix of size ($m \times n$), is defined as;

$$\boldsymbol{M_{ij}} = e^{-\frac{t_i}{\tau_j}}. \qquad 4.10$$

During the fitting process the experimental matrix needs to be convoluted with the IRF; however, as the experimental matrix is independent of the changes in the fitting argument (A-factors), this convolution only needs to be performed once (although, strictly speaking, this would not be true if the q-shift varied). An example of the convoluted experimental matrix is shown in Figure 4.1.

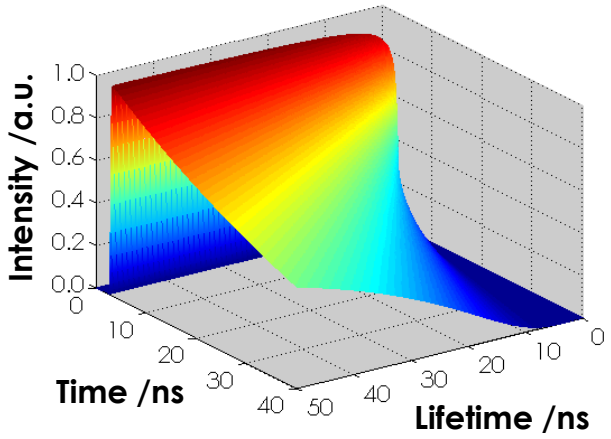

Figure 4.1 - An example of an experimental matrix, M, after convolution with a simulated IRF (a narrow Gaussian centred at t = 3 ns). The matrix was created with 201 lifetimes (τ) of between 0.01 ns and 50 ns. The reduction in attenuation rate as lifetime increases is clear to see.

One of the challenges of using distributive methods for ill-posed problems is the vast number of degrees of freedom; there are many possible solutions and it is difficult for a standard least-squares analysis to optimise to a reliable minimum. It is therefore necessary to regularise the problem by applying some additional constraint function, $C(\boldsymbol{a})$. The weighting of the constraint function is controlled by a regularisation parameter, γ, and the general mathematical description of the problem becomes of the form;

$$\dot{\boldsymbol{a}} = \arg\min_{a}\left[\frac{1}{2}\|\boldsymbol{F_N} - \boldsymbol{Ma}\|_2^2 + \gamma C(\boldsymbol{a})\right], \qquad 4.11$$

where $\dot{\boldsymbol{a}}$ is the argument of the minimum (the set of values of $\boldsymbol{a}$ that minimises the function) and $\|\boldsymbol{F_N} - \boldsymbol{Ma}\|_2^2$ is the square of the Euclidean (l_2) norm of the difference

between the decay, $\boldsymbol{F_N}$, and the fit, $\boldsymbol{Ma}$. Note that the premise of the Euclidean norm will be discussed in more detail in a later section and that the term may contain some weighting factor that is not explicitly shown in the expression above.

Maximum Entropy Method and Exponential Series Method

The maximum entropy method (MEM) has been used in a number of studies to analyse time-resolved fluorescence data.[25–28] In this case the constraint function is of the following form;

$$C_{MEM}(\boldsymbol{a}) = -\sum_{i=1}^{n} a_i - b_i - a_i \log\left(\frac{a_i}{b_i}\right), \tag{4.12}$$

where the set of values $\boldsymbol{b} = [b_1, b_2, b_3, \dots, b_n]$ represent a default model for the system. As there is no default model that describes the dynamics of fluorescence emission, b_i values are generally set to a constant value. This favours equal contribution from all lifetimes and, subsequently, means that there is limited benefit in using the additional constraint to improve the resolution of the obtained distribution.[3] Indeed, Siemiarczuk, Wagner, and Ware found little difference between the MEM and the closely related exponential series method (ESM), which completely drops the regularisation C_{MEM}, except for in the case of a single exponential (where iterative reconvolution would be the preferred method anyway).[27]

An initial attempt to replicate the ESM produced a discontinuous distribution of probability amplitudes, $p(\tau)$. This did not seem physically realistic and so it was concluded that a *smoothness* constraint should be applied. The use of smoothness was, partly, inspired by the lifetime distribution analysis used by FAST.[16] The square of the second derivative of the probability distribution of A-factors was thought to be a reasonable first approximation of smoothness, denoted here as the constraint C_{ESM}; however, in practice, the value was calculated by summing the square of the second difference between A-factor amplitudes;

$$C_{ESM}(\boldsymbol{a}) = \sum_{i=1}^{m-2} [(a_{i+2} - a_{i+1}) - (a_{i+1} - a_i)]^2. \tag{4.13}$$

Pragmatically, this strategy was successful at improving the continuity of the probability distribution; however, it would be beneficial to use a better measure of the smoothness of the distribution in future versions of the analysis code.

Compressed Sensing and Basis Pursuit Denoising

Groma *et al.* recently introduced a novel approach to analysing time-resolved fluorescence data.[3] For this method the constraint function is the l_1-norm of the vector of A-factors (Equation 4.14) and the formulation of the minimisation problem becomes that of basis pursuit denoising (BPDN);[29]

$$C_{BPDN}(\boldsymbol{a}) = \|\boldsymbol{a}\|_1 = \sum_{i=1}^{n} |a_i|. \qquad 4.14$$

The analysis process for this method can be summarised in the following way: find the simplest solution (by minimising C_{BPDN}) which can account for the experimental results (minimising the least-squares error, χ_R^2). This strategy follows the ideas of compressed sensing;[30–32] that it is possible to reconstruct the majority of a sparse (or compressible) signal by only using the most important elements. Compressed sensing has found application in, among others things, MRI,[33,34] super-resolution imaging,[35] and lensless, single-pixel cameras.[36]

To elucidate why the C_{BPDN} constraint helps to find the simplest solution, it is informative to outline the difference between l_1- and l_2-norms. The general form of the l_d-norm ($d \geq 1$ and real) of a vector of length, n, is given by;

$$\|\boldsymbol{x}\|_d = \left(\sum_{i=1}^{n} |x_i|^d \right)^{\frac{1}{d}}. \qquad 4.15$$

The norm provides a measure of the length of a vector in strictly positive terms. The Euclidean norm (where $d = 2$) is the typical calculation performed when estimating the error between the fit and true values, exemplified by the χ_R^2 statistic. Figure 4.2 shows a visual comparison of the error estimated by the l_1-norm and the l_2-norm in a 2D system.

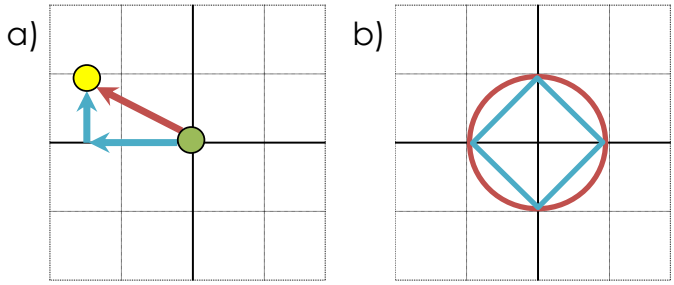

Figure 4.2 – Diagrammatic representation of l_1-norm (cyan) and l_2-norm (red) characteristics in a 2D system. a) The distance from the green dot (true value) to the yellow dot (fit value) is a visual representation of the error of the fit. Each coordinate axes represents a single component (A-factor) of the fit. The l_1-norm path length to the yellow dot from the green dot is larger, and thus more penalised, than the l_2-norm path length. b) The unit boundaries for l_1-norm and l_2-norm. The boundaries only meet when a single component is used.

In the l_1-norm case it is only possible to *travel* along one axis at a time (hence the situation is often labelled the Manhattan taxi driver problem) and results in a square unit boundary (Figure 4.2b). In contrast, the l_2 unit boundary is circular because there is no penalty for travelling along both components simultaneously. Comparison of the unit boundaries shows that the l_1-norm only matches the l_2-norm when a single component is used. In other words, for a given l_2-norm error, the l_1-norm error is minimised when there is reduction in the complexity of the solution. Addition of the C_{BPDN} constraint therefore favours simpler (sparser) solutions; that is, those with the fewest necessary components.

4.1.2.3 Simulated Annealing

Fogarty, Jones, and Camp explored an elegant approach to analysing time-resolved fluorescence data that used a simulated annealing (SA) method to deconvolve the true fluorescence decay and instrument response function from the measured signal.[6] The method was found to perform well against the typical iterative reconvolution approach used by most commercial analysis software. Similar to iterative reconvolution, the SA approach aims minimise the difference between the measured decay and an estimate of the true fluorescence decay convoluted with the measured IRF; however, unlike iterative reconvolution, the method does not impose any restriction upon the form of the guess of the true decay. The full detail of SA optimisation process is described by Fogarty, Jones, and Camp.[6] Briefly, a guess, F_G, is

made of the true fluorescence decay. The intensity of each channel of F_G is then sequentially varied (*moved*) by a random factor obtained from the range $\{e^{-\varepsilon}, e^{\varepsilon}\}$, where ε is a control parameter that is adjusted as the calculation proceeds. Each move results in a change in the least-squares error of the fit, denoted here as ΔE, and is either accepted or rejected on the basis of a comparison between a random number, r, that is generated uniformly within the range {0, 1} and a Boltzmann-inspired factor, b_f, given by

$$b_f = \min\left(1, e^{-\frac{\Delta E}{T}}\right), \qquad 4.16$$

where T is another control parameter that acts like the temperature of the system. Improvements to the fit are always accepted ($\Delta E < 0$, $b_f = 1 \geq r$) but if a move worsens the fit (a *bad* move, $\Delta E > 0$) then it is accepted if $b_f \geq r$ but is otherwise rejected; in which case the original value of F_G is restored. The reason for accepting some of the bad moves is that it avoids falling into local minima (and staying trapped there). Once all channels have been altered in a cycle the process is repeated. After a certain number of cycles the move size is decreased and the stringency for accepting a bad move is increased; in other words, ε and T are reduced. In a sense, the system begins at a high temperature, where moves are allowed to be large, and then the temperature is gradually reduced, restricting the move size. The optimisation finishes when the fit is of adequate quality or the cumulative improvement of all of the moves in a single cycle is below a certain threshold. The main benefit of this method is that subsequent fitting of the obtained fluorescence decay is unbiased by model choice. This offers the possibility of direct comparison of conventional multi-exponential analysis with non-exponential models (such as those based on distributions of decay constants). Possible non-exponential models that could be used during analysis include power laws,[37,38] stretched-exponential decays,[39] and Lorentzian[40] and Gamma[6] distributions. It should be noted that modification of Equation 4.3 could permit the use of non-exponential models for iterative reconvolution analysis; however, the additional complexity added by convoluting the model function with the IRF means that the results might be less reliable.

4.2 Methods

4.2.1 Experimental and Simulated Decays

The experimental (*real*) data used for this study was the same data collected for the previous chapter. Simulated decays were created by convoluting a known decay form with an experimental IRF, which is shown in Appendix 4, §4.1. The same IRF was used during fitting of these decays. Poisson noise was added after convolution to avoid imposing the IRF structure into the noise. The number of counts in the peak channel was typically set to around 10,000 to match the experimental data. A summary of the simulated decays is given in Table 4.1. It should be noted that it was more convenient to determine decay constants, k, rather than lifetimes, $\tau = 1/k$, during fitting.

Decay	Decay Form	Parameters
Exp1: Mono-exponential	$F_N = e^{-kt}$	k = 0.2 ns^{-1}; τ = 5 ns
Exp4: Four-exponential	$F_N = \sum_{i=1}^{4} a_i e^{-k_i t}$	**i** / **a_i** / **k_i /ns^{-1}** / **τ_i /ns** 1 / 0.60 / 20.0 / 0.05 2 / 0.20 / 2.20 / 0.45 3 / 0.15 / 0.40 / 2.50 4 / 0.05 / 0.11 / 9.09
Gamma1: Γ Distribution	$F_N = \frac{1}{(1+\kappa t)^{\alpha+1}}$	α = 20; κ = 0.01 ns^{-1} τ* = 4.55 ns (Mode) <τ> = 5.00 ns (Average)

Table 4.1 – Parameters and decay form used for simulated data.

One of the simulated decays was based on a Γ distribution of lifetimes to test the ability of the analysis techniques to handle non-exponential decays. In this case the probability for decay constant k is given by;

$$p(k) = \frac{1}{\Gamma(\alpha+1)\kappa}\left(\frac{k}{\kappa}\right)^{\alpha} e^{-\frac{k}{\kappa}}, \qquad 4.17$$

where α and κ are shape and scale parameters, respectively, and $\Gamma(x)$ is the Gamma function. An analytical summary of the properties of the Γ distribution is given by Fogarty, Jones, and Camp.[6] Of relevance to the present study is the fact that the decay form when using the Γ distribution is relatively simple and that the distribution of lifetimes has a well-defined mode and mean, which correspond to the most probable lifetime and the average lifetime, respectively (see Table 4.1).

4.2.2 Decay Analysis

All fits were performed on a standard PC. Commercial software package FAST[16] was used to provide reference fit values using the iterative reconvolution method. In addition, the distribution analysis method of FAST was also tested but it should be noted that the precise analytical process used in this case was unknown.

Fitting with other methods was performed using in-house scripts written for MATLAB.[41] A full description of the code will not be presented here but the more salient features of each analysis technique will be addressed below.

Before fitting was performed it was necessary to carry out some preparatory work. The background level of the measurements of the IRF and the fluorescence decay was estimated by using the average intensity of a set of channels (usually the first 20 channels) before the rise of the data; the background was generally less than a single count. The rise position was determined by using a threshold value, which was typically set to 0.1% of the peak channel count above the background level. Channels preceding the rise position were truncated to avoid problems during convolution. It should be noted that the fitting range was generally defined from the peak channel of the IRF to the channel of the fluorescence decay that was either around 10 times higher than the background level or equivalent to 0.1% of the counts in the peak channel, whichever was greater.

The exponential series method (ESM) optimisation used MATLAB's *fmincon* class, which finds the minimum of a constrained nonlinear multivariable function, with the *Active-Set* algorithm[1] (which was a somewhat pragmatic choice). A total of 201 fixed lifetimes that were logarithmically spaced between 0.01 ns and 50 ns were used during optimisation. A-factors were restricted to positive values between 0 and 1 and MATLAB's *MultiStart* class was used to optimise from five different starting points; four homogeneous distributions where all A-factors were initially set to the same constant ($1/n$ (where n = number of A-factors), 0, 0.5 or 1) and one distribution where A-factors were initially set to randomly generated values. A-factors were initially set (but not fixed) to zero if the corresponding lifetime was on the same order-of-magnitude as the channel width; this was done to help prevent fitting noise. The optimisation which gave the best minimisation out of the five trials was then used as the fit model. As each optimisation was independent of each other, it was possible to use parallel processing to reduce the analysis time. Typically the regularisation parameter, λ, was set to a value of 0.1 for the ESM. This value was found to be a good trade-off between improving the

smoothness (continuity) of the distribution while maintaining a reasonable resolution in the lifetime dimension, which allowed discrete lifetimes to be recovered.

The SparseLab toolbox[42] from Stanford was used for optimisation during basis pursuit denoising (BPDN, compressed sensing) analysis. Groma *et al.* investigated a number of different minimisation algorithms for BPDN but found that the one used by SparseLab (primal-dual log-barrier algorithm[43]) was best suited for fitting fluorescence decays.[3] The regularisation parameter, λ, was set to a value of 10^{-3} for BPDN analysis. This value was found to give a good balance between finding a sparse solution while avoiding oversimplification of the underlying dynamics. A total of 601 fixed lifetimes that were logarithmically spaced between 0.01 ns and 100 ns were used during optimisation. It should be noted that in the study by Groma *et al.* the number of time channels (m = 51, logarithmically spaced) was less than the number of lifetimes used during fitting (n = 412); in this situation the system is underdetermined ($n > m$).[3] For the present study the number of channels used (m ~1800, linearly spaced) was significantly greater than the number of lifetimes (n = 601); the system was overdetermined ($n < m$). Using an overdetermined system may have introduced instability into the optimisation but further investigation would be required to confirm the significance of any effects.

The core of the simulated annealing (SA) analysis code was ported to MATLAB from the Fortran version written during the parent study by Fogarty, Jones, and Camp;[6] however, some modifications were made that will now be discussed. In the original study, the initial guess of the true decay was estimated by tail-fitting (fitting from the peak of the measured decay without convolution) with a mono-exponential decay. This process was deemed superfluous because the tail of the measured decay could be used without modification. The q-shift (the relative temporal delay between the rise of the IRF and the measured decay) had been set by performing an iterative trial using the initial guess of the true decay. The best q-shift that was obtained was then fixed for the remainder of the SA process. This approach was found to give q-shift values within a channel width of those obtained by iterative reconvolution methods, which include the q-shift as a variable parameter during the full optimisation; however, small errors in the q-shift can have significant influence on the accuracy of the fit. The estimate of the q-shift was improved in the MATLAB version of the SA analysis by interpolating the IRF to allow sub-channel shifts. The q-shift was also added into the optimisation cycle to ensure that the best possible value was obtained; however, it is worth noting that the changes in q-shift were generally very small from the initial estimate. Typical initial

values used for the SA control parameters and were ε = 0.0005 and T = $E_G/(10000m)$, where E_G was the least-squares error using the initial guess of the true fluorescence decay and m was the total number of time channels. These parameter values were slightly different to those used in the original study. After 50 optimisation cycles had been performed both parameters were reduced by multiplying by a factor of 0.99. Calculations typically ran for between 10 to 20 minutes, which corresponded to thousands of optimisation cycles. As the true fluorescence decay should be of monotonically decreasing form, it was concluded that it would be beneficial to restrict the difference in intensity between neighbouring channels of the guess. This was achieved by applying a smoothing filter to the guess after a number of optimisation cycles (usually 50) had been carried out. The error weighting of the fit in the parent study was set to $1/\sqrt{D_N}$ because the short-time values were found to be extremely noisy when using the typical weighting of $1/D_N$ (which is equal to the reciprocal of the Poisson variance). This noise was not observed in the present study (perhaps due to the difference in parameters used) and so the more common $1/D_N$ weighting factor was used.

4.3 Results and Discussion

This section will provide an outline of the fits obtained using the different analytical methods for each of the decay systems that has been studied. Following this there will be a brief evaluation of the capabilities of the different analytical methods. For convenience, the notation $u\delta + v\Gamma$ will be used to describe a fit model containing u discrete exponentials and v Γ distributions. Unless specifically stated, it should be assumed that each fit was of adequate quality (low χ_R^2 value and randomly distributed residuals) to be a plausible model for the system under study. It should be noted, however, that residuals from the fits of the SA approach decreased towards longer times. This is indicative of the fact that the use of the Poisson error as a weighting factor is not appropriate for this method. Nevertheless, the fit results that were obtained using this strategy were deemed satisfactory for this preliminary study. The issue of weighting is discussed in more detail by Fogarty, Jones, and Camp.[6]

Comparison of the fits is most easily achieved by visual inspection. In the following sections, the probability amplitude for each component is plotted against the corresponding lifetime on a logarithmic scale abscissa. Discrete exponentials are represented as circular markers (note the dotted lines are physically meaningless but are presented to guide the eye) while distributions are shown as continuous lines. Amplitudes have been scaled so that the largest component has a value of one. For the mixed, $2\delta + 1\Gamma$ model the Γ distribution has been scaled such that its maximum amplitude corresponds to its overall contribution to the decay rather than scaling each individual component (the latter approach would have resulted in very small amplitudes because of the large number of contributing components). A point of note for the visualisation of the distributions is that the apparent peak width is skewed by the use of a logarithmic timescale; a similar visual distortion occurs for any apparent shift in lifetime values between analysis approaches (differences in short lifetime components are exaggerated compared to differences in long lifetimes).

A summary of the fit parameter obtained for each method is given after each plot of lifetime amplitudes. For distributive methods the full-width at half-maximum (FWHM) is provided; this value reflects the spread of lifetimes that account for the decay component. Note that FWHM values are not given in the case of the BPDN analysis because, typically, only a single value contributed to the decay component. The decay components of the ESM analysis were characterised by fitting the peaks with Gaussian functions; this required transformation of the probability amplitudes to

account for the use of logarithmically spaced lifetimes.[27] The amplitude-weighted lifetime, $\langle\tau\rangle = \sum_i^n a_i\tau_i$, is proportional to the integrated area under the associated decay curve and so is provided as a rough measure of the overall character of the calculated decay model.

Due to the diversity of the different analytical methods used during the study, it was necessary to devise a standard measure of the fit quality. The strategy that was adopted was to simulate a decay based on the obtained model parameters and then compare its convolution with the measured IRF against the original measured decay. The fit quality was then determined using a parameter, ϕ, which was based on the χ_R^2 statistic (Equation 4.7). A lower value of ϕ indicated a better fit. In the case of the simulated data it was possible to use the original input parameters to give an estimate of the best possible ϕ value; however, it should be noted that, as noise was added to the convoluted decay, it was possible for the fits to obtain a lower ϕ value than the simulated parameters. There are some other caveats to the comparison of ϕ values that should be highlighted.

Firstly, although calculated in essentially the same way, ϕ has been differentiated from χ_R^2 because the analysis methods used different minimisation conditions; calculation of ϕ may have been biased towards certain analysis strategies.

Secondly, a scaling factor had to be applied when calculating ϕ because of the normalisation of the probability amplitudes that was performed during the original fitting process. The q-shift also had to be adjusted to account for the value obtained in the original fit. In practice, the scaling factor and the q-shift were determined by a least-squares optimisation. The background level was generally fixed to a predefined value but could also have been included in the optimisation. It is worth pointing out that this secondary optimisation may have favoured particular analysis approaches, which may have distorted relative values of ϕ that were obtained.

Thirdly, the precise details of the fitting process used by FAST were unknown and so it was not possible to accurately evaluate the quality of the associated fits using the approach described above. As such, the χ_R^2 values calculated by FAST are provided for additional comparison.

Fourthly, the residuals associated with the fit to determine ϕ were not always (nor expected to be) evenly distributed across all time points because of the errors in the model parameters; this was particularly true for systems where there was a short lifetime component (which can suffer from significant error).

Finally, the number of degrees of freedom in the distribution analysis methods was difficult to quantify. For example, in the case of the BPDN analysis there were a total of 601 amplitudes that could vary; however, the final number of parameters that contributed significantly to the fit was typically less than four and, in addition, the parameters were not independent of each other. The standard approach that was adopted was to use the total number of variable parameters (for example, 601 for the BPDN analysis) for number of degrees of freedom when calculating ϕ; however, this may have led to artificially high ϕ values for the distributive methods (especially the BPDN analysis).

From the above discussion, it should be clear that ϕ values only provide a very rough comparison of the quality of the fits obtained for the different analysis methods.

4.3.1 Analysis of Simulated Decays

Figure 4.3 shows a visual representation of the fits obtained for the simulated mono-exponential decay (Exp1) and Table 4.2 shows the associated fit parameters. This example represents the simplest possible fluorescence lifetime system; the decay is defined by a single lifetime, 5 ns, that is considerably longer than the IRF width (~100 ps) but only a fraction of the time range of the measurement (~50 ns). In addition, only Poisson noise is present in the simulated decay and there is also no need to compensate for a possible q-shift. If an analysis approach is to have any standing then it must be able to accurately describe this system.

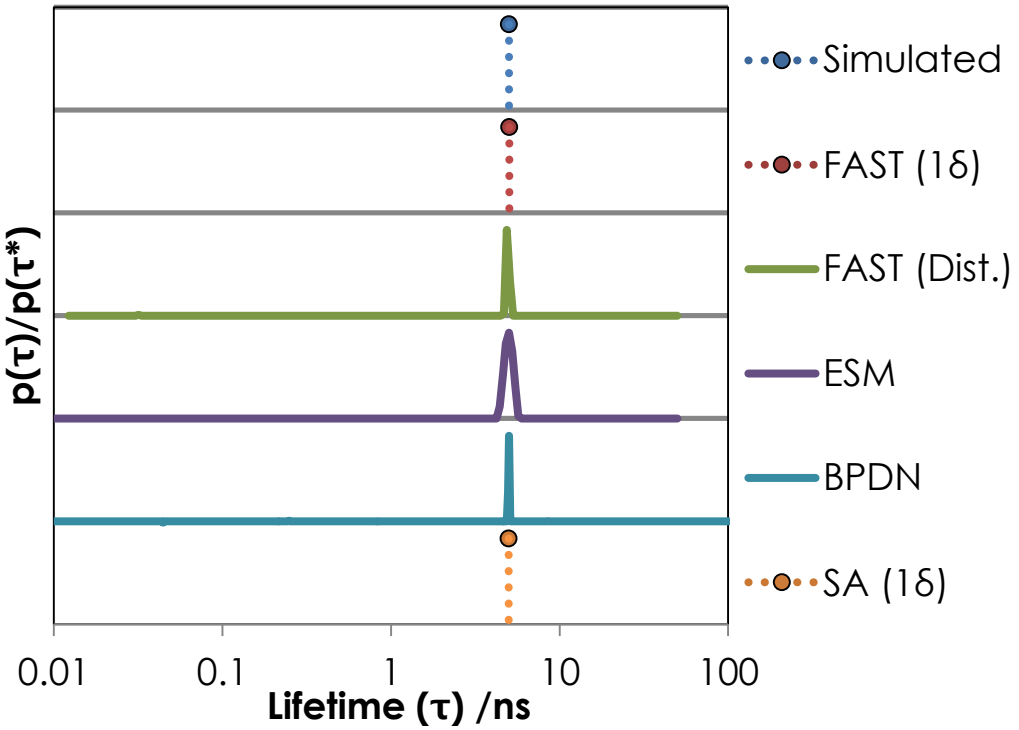

Figure 4.3 – Comparison of fits for a simulated mono-exponential decay. All methods show a single component at a lifetime close to the simulated value of 5 ns. Distributive methods have negligible probability amplitudes at other lifetime values.

Exp1	Lifetime (FWHM) /ns	ϕ (FAST χ^2_R)
Simulated	5.00	1.00
FAST (1δ)	5.03	1.07 (1.00)
FAST (Dist.)	5.03 (0.12)	1.40 (1.08)
ESM	4.98 (0.44)	1.10
BPDN	5.01	1.30
SA (1δ)	5.00	1.00

Table 4.2 – Summary of lifetimes obtained for a simulated mono-exponential decay.

Encouragingly, the results show that all methods were successful when analysing the simulated mono-exponential decay. The fitted lifetime has less than 1% error to the simulated value and ϕ is generally close to the value obtained using the simulated lifetime. Note that, as discussed above, the relatively large ϕ values for the distributive

methods are most likely an artefact of the number of degrees of freedom used. The ESM and FAST distribution fits both give fairly narrow peaks. Ideally, the distributive methods would give a single point but this is unreasonable to expect given the presence of noise and limitations of the optimisation. Note that the ESM peak could have been made narrower by decreasing the weight of the regularisation for smoothness; however, it was deemed more appropriate to use the same regularisation parameter for all of the systems studied. The influence of noise is most significant when considering times on the same order of magnitude as the channel width (~0.01 ns) and so it is reassuring to see that, despite the freedom to change, the distributive methods show negligible amplitude at short lifetime values.

Figure 4.4 shows a visual representation of the fits obtained for the simulated four-exponential decay (Exp4) and Table 4.3 shows the associated fit parameters. This example was used to determine how well the analysis methods would cope with the presence of multiple decays that were evenly distributed over a large timescale range. The lifetime values used were inspired by the 2AP-N dinucleotides studied in the previous chapter and so this simulation can be taken as an idealised version of a real system.

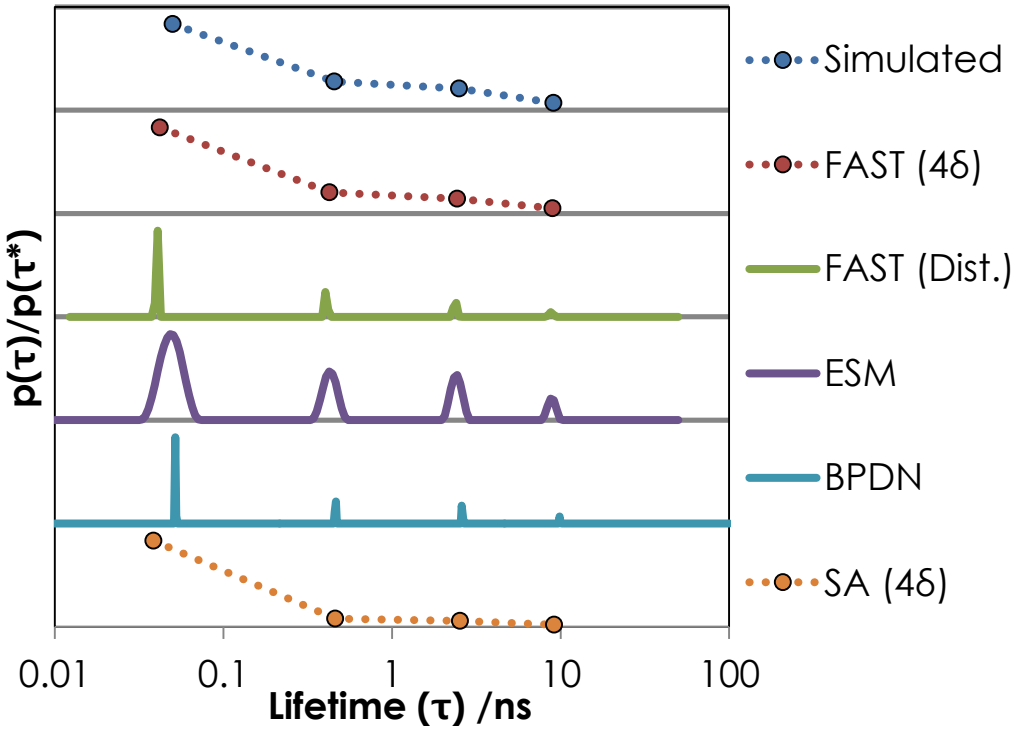

Figure 4.4 – Comparison of fits for a simulated four-exponential decay. All methods show four distinct components. With the exception of the short lifetime component of the SA 4δ fit, lifetimes and probability amplitudes are visually well-matched to the simulated decay parameters.

Exp4	Lifetime (FWHM) /ns				A-Factor				$\langle\tau\rangle$ /ns	ϕ
	τ_1	τ_2	τ_3	τ_4	A_1	A_2	A_3	A_4		$(\chi_R^2)^\dagger$
Simulated	0.05	0.45	2.50	9.09	0.60	0.20	0.15	0.05	0.95	1.07
FAST (4δ)	0.04	0.43	2.44	8.95	0.68	0.17	0.12	0.04	0.75	1.51 (1.05)
FAST (Dist.)	0.04 (0.00)	0.41 (0.01)	2.42 (0.06)	8.95 (0.25)	0.61	0.19	0.14	0.05	0.92	1.15 (1.13)
ESM	0.05 (0.01)	0.43 (0.06)	2.40 (0.26)	8.84 (0.68)	0.58	0.21	0.15	0.05	0.96	1.13
BPDN	0.05	0.46	2.59	9.85	0.59	0.21	0.16	0.05	0.99	1.35
SA (4δ)	0.04	0.46	2.54	9.14	0.84	0.08	0.06	0.02	0.39	15.02

Table 4.3 – Summary of lifetime parameters obtained for a simulated four-exponential decay. † FAST χ_R^2 values.

Generally speaking, the analysis methods perform very well at recovering the underlying model; both lifetime and probability amplitudes are well-matched. Given

the complexity of the simulated decay, it is particularly pleasing to see the success of the distributive methods, which were not restricted by any functional form during fitting. The only noteworthy discrepancy in the fits comes from the estimate of the probability amplitude of the short lifetime component; this component is overestimated by the 4δ fits of FAST and, especially, the SA method. Little can be said about the causes of the error in the FAST fit because of the *black box* nature of the program; however, the error in the SA method is likely due to an artefact of the deconvolution process. As mentioned previously, the short lifetime component is most sensitive to noise and so it is possible that the deconvolution process suffers from a similar problem. It is also conceivable that the SA input parameters that were used were not appropriate for this system; however, further investigation is required to assess this possibility. It should be noted that the very high ϕ value obtained for the SA (4δ) fit does not represent the quality of the fit to the deconvoluted decay itself, which was deemed to be adequate. Calculating a value similar to that of ϕ used here may therefore be a good strategy for checking the solutions of methods that deconvolute decays; given the error observed here, purely relying on the accuracy of the deconvolution process would seem naïve.

Figure 4.5 shows a visual representation of the fits obtained for a simulated Γ distribution of decays (Gamma1) and Table 4.4 shows the associated fit parameters. This example was created to establish whether or not the analysis methods were capable of modelling a system where there was a spread of similar lifetimes rather than well-defined, discrete lifetimes. This situation might arise, for example, in a Förster Resonance Energy Transfer (FRET) system where there is a distribution of distances between chromophore pairs.[6,12,27,44]

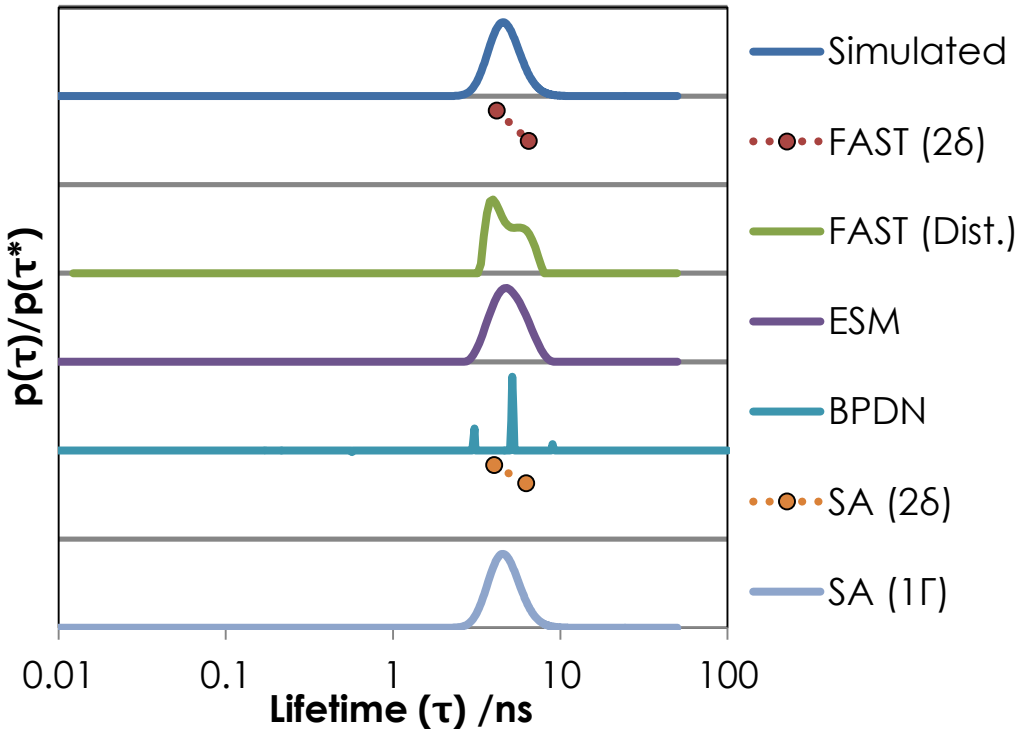

Figure 4.5 – Comparison of fits for a simulated decay based on a Γ distribution of lifetimes. All methods provide fit parameters that fall, approximately, within the limits of the simulated distribution of decay lifetimes. The fits are quite diverse between the different methods and both discrete and distributive models are able to adequately describe the simulated decay.

Gamma1	Lifetime (FWHM) /ns			A-Factor			$\langle\tau\rangle$ /ns	ϕ (FAST χ_R^2)
	τ_1	τ_2	τ_3	A_1	A_2	A_3		
Simulated	<τΓ>: 5.00 κΓ: 0.01 τΓ*: 4.55 αΓ: 20.0			AΓ: 1.00			5.00	1.00
FAST (2δ)	4.18		6.48	0.63		0.37	5.03	1.04 (1.00)
FAST (Dist.)	4.29 (0.49)		6.29 (0.66)	0.63		0.38	5.04	1.31 (1.07)
ESM		4.76 (1.47)			1.00		4.76	1.06
BPDN	3.07	5.17	8.98	0.17	0.78	0.05	4.98	1.26
SA (2δ)	4.03		6.26	0.57		0.43	4.99	0.99
SA (1Γ)	<τΓ>: 5.00 κΓ: 0.01 τΓ*: 4.54 αΓ: 20.08			AΓ: 1.00			5.00	1.00

Table 4.4 – Summary of lifetime parameters obtained for a simulated decay based on a Γ distribution of lifetimes.

The fit results of this system exemplify the difficultly in recovering the true underlying model when there is no *a priori* knowledge to discriminate against certain models: the FAST discrete exponential analysis suggests that a bi-exponential model is present; the FAST distribution analysis suggests that two narrow distributions (or perhaps a single broad distribution) of decay lifetimes is present; the ESM analysis suggests a single broad distribution of decay lifetimes is present; the BPDN analysis suggest that a three component model is present; finally, the SA methods suggest that either bi-exponential model or a single Γ distribution of decay lifetimes is present. All of these models adequately describe the measured decay and so it would be impossible to discriminate the true model by a simple comparison of the fits. One possible way to improve upon the confidence of a particular model is to simulate a decay based on the calculated parameters and then re-fit to see if a consistent set of probability amplitudes is obtained. Figure 4.6 shows a visual representation of the fits obtained for a bi-exponential decay simulated from the discrete exponential FAST fit of Gamma1. This decay simulation will be denoted as Exp2G to highlight that it came from a fit of the Gamma1 decay. Table 4.5 shows the associated fit parameters from this system. It should be noted that the spurious, short-lifetime peak in the FAST distribution fit has been omitted from the analysis and discussion presented below.

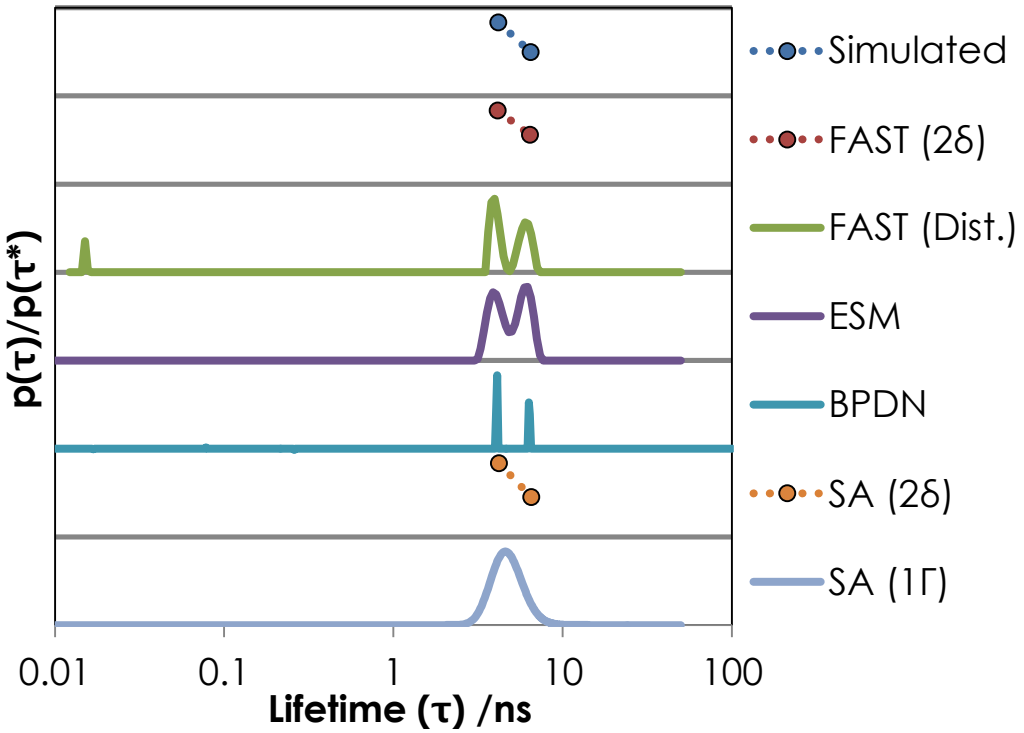

Figure 4.6 – Comparison of fits for a simulated bi-exponential decay created from the discrete FAST fit of a simulated Γ distribution of decays (Figure 4.5). All of the methods generally show two distinct components that are consistent with the simulated decay parameters; however, the SA approach also provides a fit based on a Γ distribution of decay lifetimes. The short lifetime component in the FAST distribution fit is an artefact of the fitting process and is omitted from discussion.

Exp2G	Lifetime (FWHM) /ns		A-Factor		⟨τ⟩ /ns	ϕ (FAST χ_R^2)
	τ_1	τ_2	A_1	A_2		
Simulated	4.18	6.48	0.63	0.37	5.03	0.99
FAST (2δ)	4.14	6.41	0.60	0.40	5.05	1.04 (0.98)
FAST (Dist.)	4.07 (0.25)	6.15 (0.49)	0.53	0.47	5.05	1.32 (1.05)
ESM	3.93 (0.56)	5.94 (1.01)	0.45	0.55	5.03	1.06
BPDN	4.11	6.31	0.59	0.41	5.02	1.25
SA (2δ)	4.21	6.54	0.65	0.35	5.03	0.99
SA (1Γ)	<τΓ>: 5.03 τΓ*: 4.59	κΓ: 0.01 αΓ: 20.90	AΓ: 1.00		5.03	0.99

Table 4.5 – Summary of lifetime parameters obtained for a simulated bi-exponential decay created from the discrete FAST fit of a simulated Γ distribution of decays (shown in Figure 4.5).

As expected, the results show that there is very little difference between the parameters obtained for the 2δ fits in Table 4.4 and Table 4.5; however, it is of concern that the Γ distribution adequately describes the bi-exponential decay. This is the reverse of the problem observed above and means that, under the signal-to-noise ratio used (10^2), it is impossible to discriminate between the 2δ and 1Γ models. Given that the 2δ model requires an extra fit parameter, it would be tempting to veer towards using the 1Γ model to explain both Gamma1 and Exp2G simulations, even though this would be the wrong model in the latter case.

Importantly, there is a significant difference in the results obtained for the distributive methods; all of these methods show two well-defined components rather than a single broad peak (FAST and ESM) or three discrete components (BPDN) that was observed for the Gamma1 decay. This discrepancy helps to rule out the 2δ model as a plausible description of the Gamma1 system (which is, of course, consistent with the fact that it is known to be the wrong model); if the distributions had been the same then it would not be possible to distinguish between the different models but, as they are different, these results show that the 2δ decay fit is distinct to the original Gamma1 decay. It should be noted that these observations also reduce the credibility of the 1Γ model as a description of the Exp2G simulation; however, this is only true because the 1Γ fit of Exp2G is essentially the same as the original distribution simulated for Gamma1. If they had been different (which would have to be assumed in a real system where the true decay form was not known) then the fitting procedure would have to be repeated by simulating a decay using the 1Γ fit of the Exp2G decay (that is, reverse the process carried out for the FAST (2δ) fit for the Gamma1 decay). This has (*de facto*) been done already because the Gamma1 model and the 1Γ fit of Exp2G are the same. Note that there is still no guarantee about any of the models being correct, the Exp2G simulation simply shows that a 2δ model is unlikely to be the true model for the Gamma1 system. Nevertheless, careful consideration of the results from all of the different methods would hopefully lead to the conclusion that the Gamma1 decay was due to a 1Γ model (or at least a model based on a distribution of decay lifetimes) and the Exp2G decay was due to a 2δ model.

The above discussion highlights the potential power of using probabilistic methods in combination with deterministic ones. Without imposing a model form, the distributive methods found a broad peak for the simulated Γ distribution but found a double peak for the simulated 2δ system. This shows that probabilistic methods could be used to give an unbiased estimate of the underlying model. On the other hand, the

deterministic methods provided more accurate model parameters than the probabilistic approaches when the true underlying model was used. Deterministic methods could therefore be used to refine the model that was initially determined by a probabilistic approach.

Of course, it is easy to argue for a particular model when the true model is already known. It may be useful to perform blind simulations in the future to determine the practical reliability of combining analysis methods. This being said, while simulated decays provide great insight into the different analysis strategies, the capability of an analysis methods should really be assessed with real data, which may contain unexpected distortions that are not considered in simulations.

4.3.2 Analysis of Real Decays

The results presented in the previous chapter highlighted the fact that there is still uncertainty about the underlying model that governs the fluorescence decay observed for 2AP-N dinucleotides. This system therefore stands as a good test of the proficiency of the time-resolved fluorescence analysis techniques investigated here. For the sake of brevity, only 2AP-G and 2AP-A dinucleotides will be considered. As a quick reminder of previous observations, 2AP-G required a four-exponential model while 2AP-A could be described by either a three- or four-exponential model. The less contentious 2AP-G dinucleotide results will be presented first.

Figure 4.7 shows a visual representation of the fits obtained for the fluorescence decay of 2AP-G and Table 4.6 shows the associated fit parameters. To be clear, the SA (2δ+1Γ) fit represents a mixed model where two exponential decays have been combined with a Γ distribution of decays. This model was inspired by the work of Fogarty, Jones, and Camp.[6] Using the 2δ+1Γ model, they found it was possible to obtain fits of similar quality to the typical four-exponential model but with fewer fitting parameters. The physical justification for the distribution of decay components was based on the argument that the 2AP moiety would experience a range of interactions between stacked and open conformations, which were, respectively, held responsible for the short and fast exponential components.

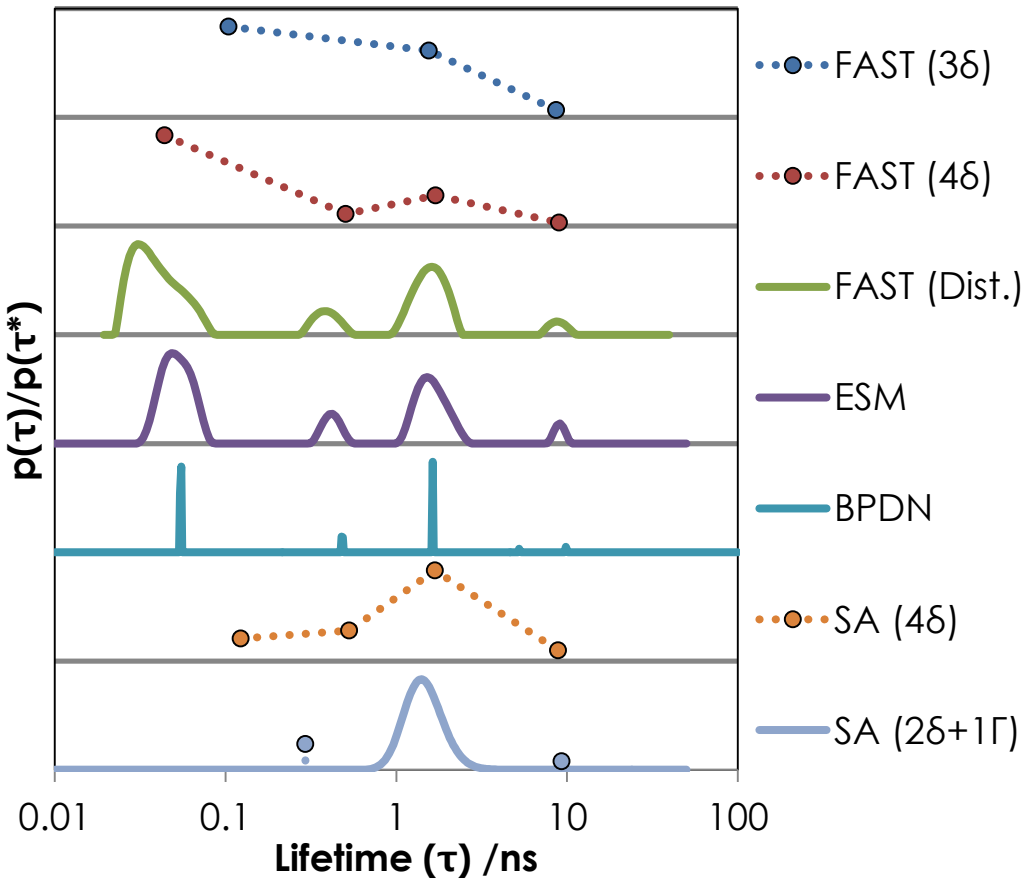

Figure 4.7 – Comparison of fits for the fluorescence decay of a 2AP-G dinucleotide (excitation and emission wavelengths were 300 nm and 370 nm, respectively). The methods mainly show four components that are consistent with the typical parameters obtained for 2AP in nucleic acid constructs. The BPDN method shows a small peak at around 5 ns; this component has likely been influenced by the presence of noise in the measured decay. The fits from the SA approach are compressed towards longer lifetimes and appear to neglect the short lifetime component observed in the other methods.

2AP-G	Lifetime (FWHM) /ns					A-Factor					ϕ
	τ_1	τ_2	τ_3	τ_4	τ_5	A_1	A_2	A_3	A_4	A_5	$(\chi_R^2)^\dagger$
FAST (3δ)	0.10	-	1.55	-	8.64	0.55	-	0.41	-	0.04	1.59 (1.35)
FAST (4δ)	0.04	0.51	1.70	-	8.98	0.66	0.09	0.22	-	0.03	2.25 (1.12)
FAST (Dist.)	0.04 (0.01)	0.40 (0.06)	1.62 (0.32)	-	9.02 (0.94)	0.56	0.09	0.32	-	0.04	1.35 (1.27)
ESM	0.05 (0.01)	0.41 (0.07)	1.55 (0.41)	-	9.02 (0.91)	0.52	0.10	0.34	-	0.04	1.29
BPDN	0.05	0.48	1.63	5.25	9.85	0.51	0.11	0.33	0.02	0.03	1.77
SA (4δ)	0.12	0.53	1.68	-	8.89	0.14	0.20	0.59	-	0.07	2.44
SA (2δ+1Γ)	0.29	$\langle\tau_\Gamma\rangle$: 1.60 τ_Γ*: 1.40		κ_Γ: 0.05 α_Γ: 13.87	9.31	0.21		A_Γ: 0.73		0.07	3.08

Table 4.6 – Summary of fit parameters obtained for the fluorescence decay of a 2AP-G dinucleotide. † FAST χ_R^2 values.

Encouragingly, the fits obtained for the FAST (4δ) model and the distributive methods are all pretty consistent with each other. This suggests that there is some standing in the use of a 4δ model to describe the 2AP-G decay, as has typically been done in previous studies.[45,46] Note that the high ϕ value for the FAST (4δ) fit is similar to that observed for the simulated Exp4 system, which was modelled quite accurately. The lower ϕ value for the FAST (3δ) fit is likely due to a bias in the ϕ calculation process and so the χ_R^2 values provided by FAST represent a more accurate description of the relative fit quality; these values suggests that the 4δ fit was adequate but the 3δ fit was not. Although the FAST (3δ) fit did not accurately describe the measured decay, this fit is shown because it highlights the fact that, in general, a deficient discrete exponential model will compensate by using a lifetime value that lies between two of the *real* components. Note that, for possible reasons discussed later, this behaviour is not observed for the 2AP-A system presented below. Despite a small (presumably spurious) peak in the BPDN analysis at around 5 ns, the similarity in the distributive analysis methods is quite remarkable considering that there has been no functional form imposed upon the fit solution. It should be noted that the short lifetime component of the BPDN distribution is composed of two neighbouring lifetime amplitudes of approximately equal weighting; this component accounts for the majority of the total decay amplitude (see Table 4.6). The consistency observed between the different distributive analysis methods greatly enhances the confidence in

the ability of these methods to accurately describe complex systems. Roughly speaking, the peaks in the distributions appear to be broader than those that were observed when fitting to the simulated decays based on discrete exponentials. This is perhaps indicative of the inhomogeneity of the local environments of the 2AP moiety in the real system; it would certainly be very surprising if all molecules exhibited exactly the same fluorescence character. Indeed, there is an argument that a discrete exponential analysis can only provide a simplistic model of a system because it cannot account for such inhomogeneity; however, in the end, this additional information might be superfluous to the requirements of the experimentalist. For example, in this case, it is probably sufficient to know the lifetimes and probability amplitudes of the four components. Of course, there may be situations where this information would be useful and so the value of a more rigorous fit should not be completely dismissed; each system needs to be considered on a case-by-case basis. Similar to the situation observed during the analysis of the Exp4 simulated decay, the SA approach appears to fail to accurately describe the short lifetime component that is implicated by the other methods; the lifetime value is overestimated while its corresponding probability amplitude is underestimated.

Figure 4.8 shows a visual representation of the fits obtained for the fluorescence decay of a 2AP-A dinucleotide and Table 4.7 shows the associated fit parameters.

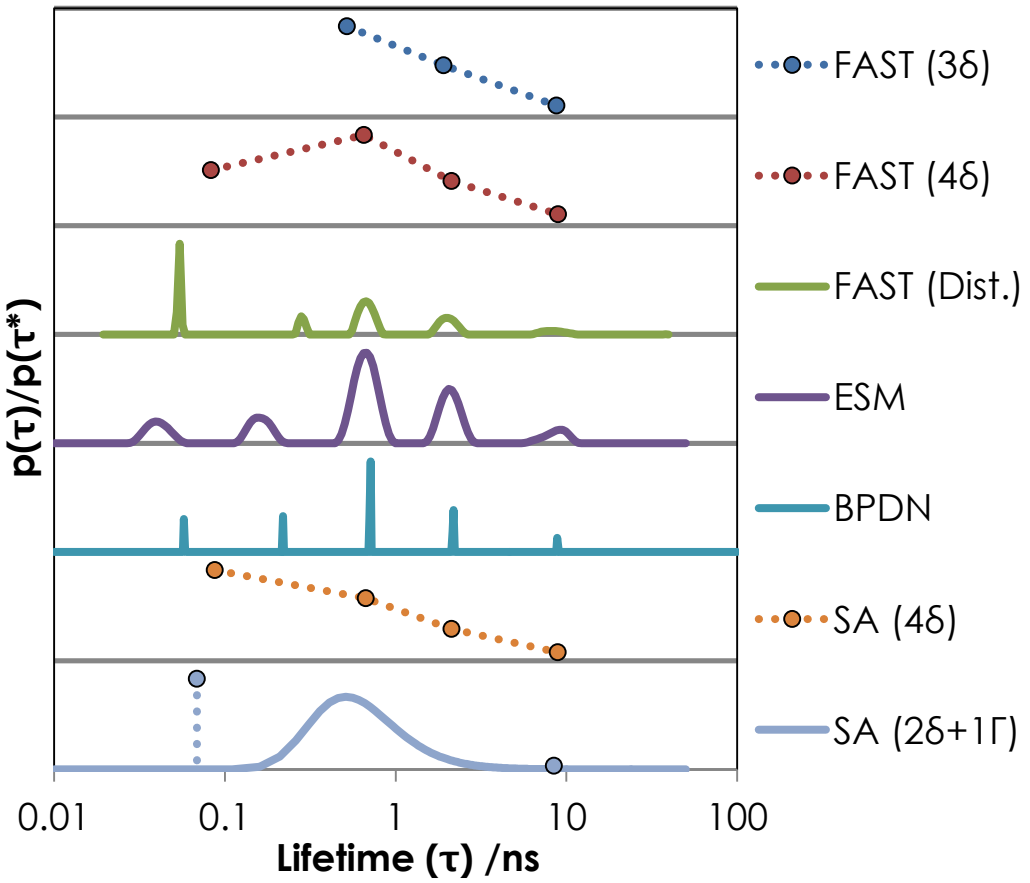

Figure 4.8 – Comparison of fits for the fluorescence decay of a 2AP-A dinucleotide (excitation and emission wavelengths were 300 nm and 370 nm, respectively). The fit parameters obtained from the different methods are quite diverse; the results suggest that three, four, or even five lifetime components may be present in the decay.

2AP-A	Lifetime (FWHM) /ns					A-Factor					ϕ
	τ_1	τ_2	τ_3	τ_4	τ_5	A_1	A_2	A_3	A_4	A_5	$(\chi_R^2)^\dagger$
FAST (3δ)	-	-	0.52	1.91	8.76	-	-	0.59	0.34	0.07	1.15 (1.12)
FAST (4δ)	0.08	-	0.65	2.13	8.97	0.27	-	0.45	0.22	0.06	1.10 (0.99)
FAST (Dist.)	0.06 (0.00)	0.29 (0.01)	0.69 (0.06)	2.07 (0.22)	8.56 (1.16)	0.20	0.08	0.40	0.24	0.08	1.19 (1.12)
ESM	0.04 (0.01)	0.16 (0.03)	0.66 (0.14)	2.06 (0.39)	8.80 (1.98)	0.11	0.13	0.46	0.24	0.07	1.13
BPDN	0.06	0.22	0.71	2.19	8.84	0.12	0.13	0.45	0.23	0.06	1.53
SA (4δ)	0.09	-	0.67	2.12	8.91	0.47	-	0.32	0.16	0.04	2.91
SA (2δ+1Γ)	0.07	<τ_Γ>: 1.40 τ_Γ*: 0.51		κ_Γ: 0.62 α_Γ: 1.15	8.47	0.44		A_Γ: 0.54		0.02	4.04

Table 4.7 – Summary of fit parameters obtained for the fluorescence decay of a 2AP-A dinucleotide. † FAST χ_R^2 values.

Interestingly, the fit results for the 2AP-A dinucleotide contrast quite considerably from those obtained for the 2AP-G dinucleotide. The most obvious difference is the fact that the distributive methods show five components rather than four. It is particularly interesting that all of the distributive approaches show the presence of five components; furthermore, the corresponding lifetimes and probability amplitudes are quite consistent between the different methods. During typical decay analysis there would be no reason to believe that a 5δ model would be appropriate to use; a 4δ fit is already powerful enough to describe (correctly or otherwise) the full behaviour of almost any measured decay. It would therefore be easy to dismiss the presence of the fifth component if it had only been in one of the distributive methods but, as it stands, it would seem that this might be a plausible description of the measured decay. It is clearly not beyond the realms of belief that there could be five underlying components within this system but the presence of the fifth component is particularly perplexing in this case because the system can, in fact, be adequately described by a 3δ model! Indeed, with the obvious exception that the 4δ fit has an extra short lifetime component, the FAST 3δ and 4δ fits are remarkably similar. At this point, it should be stressed that it is very easy to read too much into the meaning of fluorescence decay fits, especially when they are based on a single measured decay and there is no other information to corroborate any conclusions made. There are a number of factors that could lead to the observation of spurious peaks in a fit (particularly those that correspond to short lifetimes where errors are more predominant) and so a much more rigorous study is required to reveal the true nature of the underlying 2AP-A model. Nevertheless, there is perhaps some value in outlining speculative explanations for the observed behaviour. As stated previously, the 3δ fit of the 2AP-A decay does not follow the usual trend of producing a lifetime between two of the components observed from the 4δ model. For this system, the lifetime values of the 2AP-A fit are very well matched to the three longest lifetime components of the 4δ model; in other words, the short lifetime component essentially stands as an outlier and does not influence the 3δ fit. This can occur because the short lifetime component only affects a small number of initial channels in the fit. As the shortest lifetime (or two shortest lifetimes in the case of the distributions) accounts for only around 25% of the overall decay amplitude in 2AP-A, there is little penalty in using the simpler 3δ model. This contrast with the 2AP-G system where the short component accounted for ~50% of the total decay amplitude as so could not be neglected. Note that the rapid decay of a component with a lifetime of only 0.05 ns means that it is even difficult to see its presence in the fit residuals. This

line of thought would suggest that the 3δ fit is missing the short timescale behaviour of the decay. On the other hand, the low contribution of the short lifetime component means that there is likely to be greater error in its corresponding fit parameters; this might lead to anomalous effects. In the opposite manner to a deficient model fitting two of the lifetimes with a single component, it is possible that an excessively powerful model (that is, a model that has more fit parameters than are really necessary) might split a single component into two lifetimes. It is therefore plausible that the distributive methods are over-fitting noise to produce the fifth component. Note that the distributive methods may show consistency between each other because they are fitting the same decay, which contains the same noise. Alternatively, a spurious short decay component may be a manifestation of an inadequacy in the convolution process. This could happen as a result of a failure to accurately reproduce the true form of the IRF or because of a poorly optimised q-shift. The above discussion would therefore likely lead to the conclusion that the 4δ fit is probably the best model to use for this system; however, this being said, it is still slightly surprising that the BPDN method, which should minimise the number of required components, shows the presence of five lifetimes.

The results from Chapter 3 showed that 2AP-A exhibits unusual behaviour compared to the other 2AP-containing dinucleotides. One of the potential causes for this apparent discrepancy was the use of an inappropriate model during fitting. It is therefore of great interest that all of the different analytical approaches essentially show the same behaviour for this system; the component associated with a lifetime of around 0.65 ns dominates the decay profile. This contrasts the behaviour of the other dinucleotides where the equivalent component (denoted A_2 (τ_2) in Chapter 3) has only minor contribution to the overall decay. On the basis of reduced charge transfer efficiency, it was reasoned in Chapter 3 that the dominance of the intermediate component in 2AP-A might be due to the fact that the component really comprises both of the short components observed for the other systems. This could still be a valid argument but it is slightly curious that the distribution analysis methods, which show two short lived components, still maintain such a dominant intermediate component. From these observations it would seem that 2AP-A really is unusual.

As stated before, a far more rigorous study is required to ascertain the most reliable description of this system. An obvious starting point would be to further investigate the 2AP-A fluorescence decays that were measured in different alcohol-water contents; observing trends might again provide better insight than

individual measurements. The important aspect to take away from this study is that the use of different analytical methods has potentially offered new insights into this complex system.

It should be noted that the results of the SA method have not been included in the above discussion because the short lifetime component is, again, poorly described; the lifetime value is consistent with the other methods but the amplitude is greatly overestimated. The longer lifetimes are fairly well-matched with the results of the other analysis methods but this is little consolation considering the interdependent nature of the fit parameters.

Now that a variety of different systems have been analysed, a brief review of the benefits and limitation of the various analysis methods that used during this study will be presented.

4.3.3 Evaluation of Time-Resolved Fluorescence Analysis Methods

4.3.3.1 FAST (Discrete Exponential and Distribution)

As expected from a commercial program, the performance of the FAST discrete exponential analysis was of very high standard. Lifetimes and corresponding probability amplitudes were generally very consistent with the simulated decay parameters. One of the limitations of using a commercial program is that only specific models are offered, none of which may be appropriate for the system under study. This being said, it would be fairly easy to add in more models to future versions of the program if it was shown that there was significant value in doing so.

Without knowing the precise mechanics behind the fitting algorithm, it is difficult to say much about the FAST distribution analysis beyond the fact that it was generally found to give plausible results that matched well with other methods. It would seem sensible to at least use this analysis method as an initial test of the underlying model.

4.3.3.2 Exponential Series Method (ESM)

Given the simplicity of the constraint function used, the results obtained from the ESM analysis are quite remarkable. Without imposing any model form on the fit the ESM analysis gave lifetime distributions that were in very good agreement with the simulated parameters. Of course, there were some limitations to the method, such as the fact that the discrete exponential components were described by peaks rather than points. Additionally, although the ESM method correctly predicted two components for the Exp2G decay, the relative weighting of the components had considerable error. This

is hardly surprising though; in addition to the influence of noise, the lifetime components were close together and the corresponding amplitudes were also fairly similar. The Exp2G system was, in effect, not well suited to probabilistic analysis. In this situation the deterministic approaches are preferred and, indeed, were found to give more accurate fit parameters.

It is worth pointing out that it was only possible to estimate the optimal value for the regularisation parameter, γ, for the ESM (and BPDN) analysis because of the use of simulated decays, which allowed direct comparison of the fits with the known form of the true decay. In future studies it might be necessary to simulate decays with similar system properties as the real data that is to be fitted (for example, number of channels and signal-to-noise ratio) to provide an estimate for the appropriate regularisation parameter weighting to use.

4.3.3.3 Basis Pursuit Denoising (BPDN)

The BPDN method appears to be an extremely proficient method for recovering a good estimate for the underlying decay model without imposing any restriction upon the form of the solution. A particularly attractive feature of the BPDN approach is the fact that it attempts to find the solution with fewest components; the simplest model. Again, there are some limits to this method that should be highlighted. As with iterative reconvolution, adequately describing the observed behaviour with a set of discrete lifetimes does not necessarily mean that this is the true model of the system. Indeed, when performing fits of decays that are known to be due to a distribution of lifetimes (for example, Gamma1) it is still possible for a discrete set of lifetimes to successfully model the data within the resolution and noise limits offered by the experimental technique. As it aims to minimise the number of contributing components, the BPDN analysis approach is inherently poor at recovering a distribution of lifetimes; however, it is encouraging that the general shape of the lifetime amplitudes was consistent with the simulated decay lifetime distribution for the Gamma1 system. Additionally, Groma *et al.* found that it was possible to recover a distribution of decay lifetimes using BPDN if a large number (~100) of decays were simulated.[3] As BPDN is sensitive to noise, each decay fit resulted in slightly different peak positions which, on average, recovered the underlying distribution of lifetimes. This is a viable process for simulations, where many decays can be generated instantaneously, but it is likely to be unfeasible for a real system due to the considerable time it takes to measure even a single decay. More of a concern, however, is the fact that the specific noise present the decay has influence on

the obtained lifetime parameters and, occasionally, the presence of extraneous components. Similar to the ESM, it would seem preferable to refine the solution obtained by BPDN by using a deterministic approach.

4.3.3.4 Simulated Annealing (SA)

The SA approach holds great potential for the analysis of time-resolved fluorescence decays because of the ability to test a wide variety of different (and perhaps more complicated) models. This deconvolution approach was found to perform well in situations where there was no short lifetime component present in the system; however, in this situation it would be likely that tail-fitting would be able to give similar results without the requirement of any convolution at all. Indeed, tail-fitting the Exp1 simulation gives a lifetime value of 5.03 ns, which is almost identical to the true value. The main benefit of any deconvolution method should be to improve upon the recovery of short lifetime components but, in this study at least, the SA method appears to distort the fluorescence decay on short timescales, which had a significant effect on the resulting fit parameters. This issue needs to be addressed if the full potential of the SA method is to be realised. It should be noted that there is some degree of freedom in the SA deconvolution process and it is possible that the parameters used for this study were simply not appropriate; however, a much more extensive study would be required to assess this possibility.

4.4 Conclusions

Various analytical methods for time-resolved fluorescence data have been investigated with real and simulated decays to try and obtain an overview of their benefits and limitations. The results have provided insight into the relative merits of using deterministic approaches, such as the commonly used iterative reconvolution method, and probabilistic approaches, such as the exponential series method (ESM) and recently proposed basis pursuit denoising (BPDN, compressed sensing) method. A deconvolution approach, based on the process of simulated annealing (SA), has also been investigated.

In addition to idealised discrete exponentials models, which are typically used to simulate fluorescence decay curves, the analytical methods were also tested with a decay that was based on a distribution of lifetimes. Such a distributive model can be more physically appropriate for some real systems and therefore allowed a more complete assessment of the capabilities of the various methods to be carried out. For the most part, all of the analytical approaches were able to recover the underlying model that had been simulated. The only major discrepancy was in the short lifetime behaviour that was obtained by the SA approach; however, as has been stated before, this may have been due to the use of inappropriate parameters during the deconvolution process. Some other limitations of the different methods have been outlined in the previous section and will not be repeated here. It is enough to note that the main outcome of this evaluation is that no single method is preferred in all situations and there is likely to be value in using a combination of multiple methods when there is ambiguity in the interpretation of the results. While this strategy still might not be enough to provide a definitive model of the system, as was apparent in the study of the real fluorescence decays of 2AP-containing dinucleotides, it would hopefully be able to provide a model that it sufficient for the purposes of the experimentalist.

It is worth pointing out that the code that was written to perform the analysis presented here is still work in progress. The following section will briefly outlined some of the potential improvements that could be made.

4.4.1 Improvements to Analysis Methods

To save having to continually convolute the large experimental matrix with a shifting IRF, the q-shift was fixed for the distributive ESM and BPDN analysis methods. To reduce the possible error this caused, the q-shift value was estimated by iterative

reconvolution of the tail of the measured decay (approximating a guess of the true decay) with the IRF; this is the same process that was used to initially estimate the q-shift during the SA analysis, as described in §4.2.2. This procedure was found to give values very close to those obtained from the FAST program after full optimisation. Nevertheless, the use of a fixed q-shift is an obvious shortfall and needs to be addressed.

Fitting of the deconvoluted decays recovered by the SA approach is another area which needs to be improved. As stated before, the (residual) weighting that is currently used is not appropriate. A more pressing issue, however, is the distortion SA appears to cause at short timescale; the fit will never produce reliable results if this problem is not addressed first. One of the challenges of improving the SA approach is the fact that the deconvolution process takes quite a long time (~10-20 minutes), which is due to the probabilistic nature of the optimisation. To put the analysis time into context, the other analysis methods generally took under one minute to finish. Refining the SA deconvolution accuracy is likely to be quite a time-consuming process.

The restriction of FAST (and other commercial analysis programs) to particular decay models inspired the development of analysis code that could perform iterative reconvolution analysis using any model of choice. Although essentially completed, the code will require rigorous testing to evaluate the accuracy of the fitting process. It will particularly interesting to see if data can be successfully fitted using more esoteric, non-exponential functions like the Gamma distribution.

In addition to the suggestions provide above, there are some more general strategies that could help to improve the fit results obtained from the various analysis methods. Clearly, the reliability of the results obtained from any analysis methods relies heavily on the quality of the data that is used. Collecting decays with extremely good signal-to-noise ratios is likely to be necessary to be able to discriminate between some of the more complex models; such as those that were required for the 2AP-containing dinucleotides decays. Being able to perform global analysis on multiple decays simultaneously would also provide a significant improvement to the robustness of the solutions obtained. FAST already has the ability to perform global analysis for its iterative reconvolution and distributive methods; it would seem quite plausible that this powerful analysis strategy could be introduced into the other methods as well.

4.4.2 Outlook

Unfortunately, despite advances in experimental equipment (higher sensitivity detectors, homogeneous wavelength response, lower background counts, and so on), the fundamental issue of obtaining the most physically realistic model of a system remains because it is not a technological phenomenon. As has been stated previously, the main problem with using the null hypothesis to determine the model which best describes the system is that it cannot definitively provide the correct model; it can only indicate that the tested model may be correct or may be wrong. Indeed, James and Ware give serious warnings over the interpretation of TCSPC data when there is not any supplementary information to corroborate the model chosen, particularly in instances where decays are collected to an inadequate signal-to-noise ratio.[47] Given the unavoidable nature of this problem, all that can be hoped for is that the model that is finally selected is the best possible choice given the information available. The results presented in this chapter have shown that the combination of probabilistic and deterministic analysis methods can potentially allow a much more confident prediction to be made about the true model that underlies the observed fluorescence decay. Giurleo and Talaga have previously discussed the potential value of this strategy whilst also noting that the immense computational power that has been afforded to researchers over the past few decades needs to be better exploited for these means.[1] Time-resolved fluorescence techniques provide crucial insight into the underlying mechanics of a great variety of different systems; it is therefore perhaps about time that the analysis methods catch-up with the experimental advances that have been made.

4.5 References

1. J. T. Giurleo and D. S. Talaga, *J. Chem. Phys.*, 2008, **128**, 114114.

2. G. Landl, T. Langthaler, H. W. Englt, and H. F. Kauffmann, *J. Comput. Phys.*, 1991, **95**, 1–28.

3. G. I. Groma, Z. Heiner, A. Makai, and F. Sarlós, *RSC Adv.*, 2012, **2**, 11481.

4. A. E. McKinnon, A. G. Szabo, and D. R. Miller, *J. Phys. Chem.*, 1977, **81**, 1564–70.

5. A. E. W. Knight and B. K. Selinger, *Spectrochim. Acta Part A Mol. Spectrosc.*, 1971, **27**, 1223–34.

6. A. C. Fogarty, A. C. Jones, and P. J. Camp, *Phys. Chem. Chem. Phys.*, 2011, **13**, 3819–30.

7. Y. Chen, B. Liu, H. Yu, and M. D. Barkley, *J. Am. Chem. Soc.*, 1996, **7863**, 9271–8.

8. B. Liu, R. K. Thalji, P. D. Adams, F. R. Fronczek, M. L. McLaughlin, and M. D. Barkley, *J. Am. Chem. Soc.*, 2002, **124**, 13329–38.

9. P. Rai, T. D. Cole, E. Thompson, D. P. Millar, and S. Linn, *Nucleic Acids Res.*, 2003, **31**, 2323–32.

10. J. M. Jean and K. B. Hall, *Biochemistry*, 2004, **43**, 10277–84.

11. R. Luchowski, Z. Gryczynski, P. Sarkar, J. Borejdo, M. Szabelski, P. Kapusta, and I. Gryczynski, *Rev. Sci. Instrum.*, 2009, **80**, 033109.

12. J. R. Lackowicz, *Principles of Fluorescence Spectroscopy*, Springer Science & Business Media, LLC, 3rd edn., 2006.

13. J. Večeř, A. A. Kowalczyk, L. Davenport, and R. E. Dale, *Rev. Sci. Instrum.*, 1993, **64**, 3413–24.

14. J. Večeř, A. A. Kowalczyk, and R. E. Dale, *Rev. Sci. Instrum.*, 1993, **64**, 3403–12.

15. W. R. Ware, L. J. Doemeny, and T. L. Nemzek, *J. Phys. Chem.*, 1973, **4561**, 2038–48.

16. FAST, *Version 3.4.1*, Edinburgh Instruments Ltd., Kirkton Campus, Livingston, United Kingdom, See; http://www.edinburghphotonics.com/, 2013.

17. PicoQuant, *J. Fluoresc.*, 2003, **13**, 289.

18. J. R. Knutson, J. M. Beechem, and L. Brand, *Chem. Phys. Lett.*, 1983, **102**, 501–7.

19. J. M. Beechem and L. Brand, *Photochem. Photobiol.*, 1986, **44**, 323–9.

20. I. H. M. van Stokkum, D. S. Larsen, and R. van Grondelle, *Biochim. Biophys. Acta*, 2004, **1657**, 82–104.

21. A. Arcioni and C. Zannoni, *Chem. Phys.*, 1984, **88**, 113–28.

22. J. M. Beechem, M. Ameloot, and L. Brand, *Instrum. Sci. Technol.*, 1985, **14**, 379–402.

23. B. Robotham, K. A. Lastman, S. J. Langford, and K. P. Ghiggino, *J. Photochem. Photobiol. A Chem.*, 2013, **251**, 167–74.

24. C. Ruckebusch, M. Sliwa, P. Pernot, A. de Juan, and R. Tauler, *J. Photochem. Photobiol. C Photochem. Rev.*, 2012, **13**, 1–27.

25. A. K. Livesey and J. C. Brochon, *Biophys. J.*, 1987, **52**, 693–706.

26. A. Siemiarczuk and W. R. Ware, *J. Phys. Chem.*, 1989, **93**, 7609–18.

27. A. Siemiarczuk, B. D. Wagner, and W. R. Ware, *J. Phys. Chem.*, 1990, **94**, 1661–6.

28. J.-C. Brochon, *Methods Enzymol.*, 1994, **240**, 262–311.

29. P. R. Gill, A. Wang, and A. Molnar, *Trans. Signal Process. IEEE*, 2011, **59**, 4595–605.

30. E. J. Candès, J. K. Romberg, and T. Tao, *Commun. Pure Appl. Math.*, 2006, **LIX**, 1207–23.

31. D. L. Donoho, *IEEE Trans. Inf. Theory*, 2006, **52**, 1289–306.

32. E. Candès and M. B. Wakin, *Signal Process. Mag. IEEE*, 2008, **25**, 21–30.

33. M. Lustig, D. Donoho, and J. M. Pauly, *Magn. Reson. Med.*, 2007, **58**, 1182–95.

34. M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, *Signal Process. Mag. IEEE*, 2008, 72–82.

35. H. P. Babcock, J. R. Moffitt, Y. Cao, and X. Zhuang, *Opt. Express*, 2013, **21**, 28583–96.

36. G. Huang, H. Jiang, K. Matthews, and P. Wilford, *20th IEEE Int. Conf. Image Process.*, 2013, 2101–5.

37. H. C. Fogedby and V. Poutkaradze, *Phys. Rev. E*, 2002, **66**, 021103.

38. J. Włodarczyk and B. Kierdaszuk, *Biophys. J.*, 2003, **85**, 589–98.

39. M. N. Berberan-Santos, E. N. Bodunov, and B. Valeur, *Chem. Phys.*, 2005, **315**, 171–82.

40. S. Bharill, P. Sarkar, J. D. Ballin, I. Gryczynski, G. M. Wilson, and Z. Gryczynski, *Anal. Biochem.*, 2008, **377**, 141–9.

41. MATLAB, *Version 8.2.0 (R2013b)*, The MathWorks, Inc., Natick, Massachusetts, United States of America, 2013.

42. D. Donoho, I. Drori, V. Stodden, Y. Tsaig, and M. Shahram, *Version 2.0*, See; http://sparselab.stanford.edu/, 2007.

43. S. S. Chen, D. L. Donoho, and M. A. Saunders, *SIAM J. Sci. Comput.*, 1998, **20**, 33–61.

44. I. V. Gopich and A. Szabo, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 7747–52.

45. O. J. G. Somsen, L. B. Keukens, M. N. de Keijzer, A. van Hoek, and H. van Amerongen, *Chemphyschem*, 2005, **6**, 1622–7.

46. X. Wu, PhD Thesis, University of Edinburgh, 2012.

47. D. R. James and W. R. Ware, *Chem. Phys. Lett.*, 1985, **120**, 455–9.

Chapter 5: A Computational Study of 2AP-Containing Dinucleotides

We are reaching the stage where the problems we must solve are going to be insoluble without computers. I do not fear computers. I fear the lack of them.

– Isaac Asimov

The main computers that run the whole thing started overwriting their own hard drives and shut down completely and the station was basically unable to control which way it was even pointing. Fortunately we were docked to the shuttle, so we could take over until we were able to replace the hard drives.

– Commander Chris Hadfield (NASA astronaut, ISS)

5.1 Introduction

This chapter reports an investigation into the conformational properties of 2AP-containing dinucleotides. As was highlighted in Chapter 2, there is remarkable similarity between the photophysical behaviour of 2AP incorporated into dinucleotide and duplex constructs. Discussion in previous chapters alluded to the fact that conformation seems to play an important role in determining the relaxation dynamics of the excited 2AP fluorophore in these systems. As a brief reminder, the shortest lifetime (τ_1, < 100 ps) is thought to be due to a conformation where rapid charge transfer occurs between efficiently stacked bases. The longest lifetime (τ_4, ~9 ns) is commonly attributed to an open (or base-flipped) conformation where the 2AP moiety does not experience significant intramolecular interaction and essentially behaves like free 2AP. The precise origin of the two intermediate lifetimes (~0.5 ns and ~2 ns) is still unknown but it is possible that they are due to intermediate conformations between fully stacked and open forms. Given the apparent disparity of the conformational constraints within the two systems, it is slightly surprising that dinucleotide and duplex systems exhibit such similar photophysical behaviour. The typical assumption of this observation is that the dinucleotide system accesses duplex-like conformations with similar probability; however, there is little direct evidence to justify this conclusion. In particular, it is not immediately obvious that a dinucleotide, in solution, would be able to maintain a stable, stacked conformation that could facilitate charge transfer between the bases. On the other hand, it is also difficult to intuitively grasp the stability of an open dinucleotide conformation, especially when there would likely be a significant hydrophobic driving force promoting closure.

Some evidence of the presence of stacking interactions in dinucleotides was provided by early circular dichroism (CD)[1–3] and nuclear magnetic resonance spectroscopy (NMR)[4] experiments. NMR results also suggested that, at 20°C, around 30% of a dinucleotide population is stacked in solution.[5] More recent time-resolved IR spectroscopy experiments have reinforced this conclusion as well as indicating that dinucleotides can, in fact, adopt conformations where interbase charge transfer is possible.[6,7]

Despite providing useful insight, the aforementioned studies do not provide the required conformational detail to understand the intricacies of the photophysics of 2AP in nucleic acids; for instance, there may be a number of conformations that could be considered to be *stacked* but each may have different photophysical properties.

Fortunately, the rapid advance of computational methods has offered a new way to interrogate the conformational stability of almost any conceivable system. The applicability of computational methods to nucleic acid structures has not been lost on the scientific community and a great body of work has accumulated in recent years. The following section will briefly discuss some of the more relevant studies that have been performed.

5.1.1 A Computational Approach to DNA Structure

Previous computational studies that have looked at base-stacking interactions have typically used non-covalently bound dimers and simply omitted the sugar-phosphate backbone. It has been common to fix the bases in canonical B-DNA conformation. This approach has provided many useful insights into the properties of nucleobases when in a DNA-like environment; however, it does not allow the differences between dinucleotide and duplex constructs to be revealed and also does not reflect the conformational mobility of DNA.

Another approach has been to scan the potential energy surface of the dimer system by choosing a particular coordinate of interest (such as twist angle, rise distance or translational displacement) and iteratively adjusting it through a range of values. In this case the bases in the dimer can be either completely frozen (single-point calculations) or allowed to optimise under the restriction of the fixed coordinate being interrogated. This method has been useful in determining an energy landscape for the stacking interaction between isolated nucleobase dimers but these studies do not account for the constraints imposed by the sugar-phosphate backbone (which will alter this landscape). Additionally, most of these dimer conformations will not be accessible when the two nucleobases are covalently bound with the sugar-phosphate backbone and so one has to question their biological relevance.

Šponer, Jurečka, and Hobza provide an in-depth review of quantum mechanical (QM) studies that have investigated base-stacking and base-pairing interactions.[8] Although mainly based on observations of dimer systems, there are a number of significant outcomes can be taken from this appraisal.

Firstly, base-stacking properties are controlled by three main contributions; dispersion attraction, short-range exchange repulsion, and electrostatic interactions. Dispersion attraction is proportional to the geometric overlap of the bases and is primarily responsible for the stability of the base-stacked structure. Equilibrium of exchange repulsion and dispersion attraction determines the rise (vertical separation)

between the bases. Electrostatic interactions are mainly responsible for the relative orientation (twist, tilt, and roll) and displacement (slide and shift) of the bases. Solvation, particularly in polar solvents, is therefore predicted to destabilise conformations optimised in the gas-phase. Interestingly *ab initio* stacking energies could be well-predicted with a simple empirical model which contained only Lennard-Jones and Coulombic terms.

Secondly, nucleobases exhibit asymmetric pyramidalisation of their NH_2 (amino/aniline) group when isolated in the gas-phase. Non-planarity is greater for the amino group of guanine compared to the aniline groups of adenine and cytosine. Asymmetry is caused by interaction with nearby exocyclic atoms. In a base-paired system planarity is restored to optimise H-bonding efficiency.

Thirdly, the stacking observed in nucleic acids does not correspond to the potential energy minimum of isolated stacked dimer structures. Scanning the potential energy surface with single-point calculations is preferred over gradient optimisation methods when searching for the global minimum of dimers as the latter can get stuck in one of many local minima present in these systems; however, single-point calculations are inappropriate for dinucleotides because of the influence of changes in the sugar-phosphate backbone as well as base-stacking.

Fourthly, the strength of the interaction between two bases within a nucleic acid construct is a product of many different factors and cannot simply be accounted for by intrinsic base-base interactions only. For instance, interactions with other proximal bases, solvent molecules, and ions (particularly heavy ions that are sometimes used to enable crystallisation) will also influence the observed properties of the specific base interaction of interest. This makes comparison between calculations and practical experiments difficult. Without ensuring the system has been correctly modelled, specific contributions may cause a large discrepancy between calculated and observed energies.

Fifthly, care must be taken when using X-ray crystal structures as a basis for an input geometry for a calculation. This is particularly true if single-point calculations are to be performed on fixed conformations. Small errors in the separation of the bases can lead to large discrepancies in energy. Perturbation of the nucleobase geometry can also cause changes in the electronic structure.

Sixthly, QM and molecular dynamics (MD) studies suggest that gas-phase nucleobase experiments (such as, Abo-Riziq *et al.*[9] and Nir, Kleinermanns, and de Vries[10]) contain a mixture of various conformational states. The ensemble of structures

is likely to include H-bonded base pairs, T-shaped dimers, and base-stacked dimers. The complex nature of these systems means that it is very difficult to separate out factors which contribute to bulk properties and so it is challenging to obtain a meaningful comparison between calculation and practical experiment.

Finally, the effect of solvent is an important consideration when performing calculations on nucleic acid constructs. A study by Kabeláč, Ryjáček, and Hobza found that the inclusion of only a handful (two to six) of explicit water molecules was enough to favour base-stacking over base-pairing.[11] Also, Fonseca Guerra *et al.* found that inclusion of water and counter-ions was required to emulate experimental base-pairing geometry.[12] Unfortunately, inclusion of explicit solvent molecules increases the computational demands of a calculation. Additionally, using only a few solvent molecules does not provide an accurate description of the effect of solvation because solvent-solvent interactions are underestimated. Šponer, Jurečka, and Hobza[8] suggest implicit models, such as the polarizable continuum model (PCM), can provide reasonable results. Although solvation energy was determined to be poorly modelled (resulting in inaccuracies in estimated binding energies) the changes in electronic structure and the wavefunction (used to calculate many intrinsic properties) could be successfully predicted.

Hunter and van Mourik investigated base-stacking in U|U and T|T dimer structures at the M06-2X/6-31+G(d) level of theory in solution and the gas-phase.[13] The use of the M06-2X functional was motivated by the fact that it had been proven to give accurate results in dispersion-dominated systems. Calculated stacking interaction energies showed excellent agreement with benchmark estimates using CCSD(T)/complete basis set. The success of the Hunter and van Mourik study is the principal reason why structures reported in this chapter were also calculated using the M06-2X functional; however, it is also worth highlighting that the protocol used is similar to that suggested by Churchill and Wetmore for a dinucleotide system.[14]

Hydrogen bonding in DNA base pairs was investigated by Fonseca Guerra *et al.*[12] This study is of note because it explicitly considered the effect of the sugar-phosphate backbone in a calculation of a Watson-Crick complex of two deoxyadenylyl-3′,5′-deoxyuridine dinucleotides. Unfortunately the results of this calculation are not discussed in much detail beyond the fact that the sugar-phosphate backbone is concluded to have little influence on base-pairing. Recent studies have emphasised the importance of the backbone structure in promoting nucleic acids into typical double-helical conformation.[15–17]

Jean and Hall performed an extensive study on the electronic structure of isolated 2AP in water and the gas-phase.[18] While the configuration interaction singles[e] (CIS) method used overestimated absolute excitation energies, the excited-state energy gaps and the direction of transition dipole moments were in good agreement with experiment. Importantly, as the results were physically realistic, this justified the use of such theoretical methods in subsequent studies which considered more complex systems.

To improve upon the accuracy of calculated excitation energies, Jean and Hall moved to using time-dependent density functional theory (TD-DFT) with the B3LYP exchange-correlation function (TD-B3LYP) in two studies of nucleobase dimers[19] and trimers[20] containing 2AP. Base-stacking in 2AP|Y dimer structures was predicted to result in the creation of an additional excited state not found in the 2AP monomer. Importantly this state was the lowest energy excited state and also had weak oscillator strength (which is a property related to the transition probability). Vibrational relaxation into this low-lying dark state was predicted to be responsible for fluorescence quenching in these systems. Interestingly, the electronic structure of 2AP|T was found to be different when fixed in A-DNA form and this was attributed to the fact that the (π-orbital) overlap of six-membered rings was greater than in the B-DNA conformation. In the A-DNA form the weak oscillator strength of nearly degenerate $S_0 \rightarrow S_1$ and $S_0 \rightarrow S_2$ transitions was predicted to account for fluorescence quenching. A similar explanation was presented for 2AP|purine dimer structures; electron delocalisation in the ground state was concluded to weaken the oscillator strength of $S_0 \rightarrow S_1$ and $S_0 \rightarrow S_2$ transitions. Calculations also found states in these systems that had considerable charge transfer character. Nonradiative relaxation into these charge transfer states (which are unavailable to the 2AP monomer) was proposed to be possible alternative quenching mechanism in base-stacked systems. 2AP-containing trimer constructs were found to undergo similar relaxation processes that were present in the dimers. Importantly, the results obtained from these investigations showed that base-stacking significantly alters the electronic structure of 2AP.

Hardman and Thompson investigated the influence of base-pairing and base-stacking on the fluorescence properties of 2AP using CIS/6-31+G(d,p) in vacuo.[21] They found that base-pairing had a negligible effect on electronic structure and

[e] Only single excitations considered.

predicted 2AP·T should essentially behave as the free 2AP molecule. In contrast, base-stacking caused perturbation to the orbitals associated with the fluorescence transition through mixing with the proximal base. The resulting reduction in the oscillator strength of the transition (by as much as 53%) was expected to cause a reduction in quantum yield. It was pointed out (under the condition that there were no other changes in the system) that this would also cause a reduction in the radiative rate constant (k_R), which would be observable as an increase in fluorescence lifetime of the 2AP. As a decrease in fluorescence lifetime is experimentally observed it was concluded that stacking must additionally cause an increase in the nonradiative rate constant (k_{NR}). The results could not rule out charge transfer as a possible source of this increase in k_{NR}. Another nonradiative pathway that was proposed was internal conversion to a dark S_2 state. When 2AP was sandwiched between two guanine bases the gap between S_1 and S_2 states decreased and so the likelihood of this transition increased.

Hardman and Thompson warn against the prediction of low-lying charge transfer states by TD-B3LYP. Due to poor reproduction of long-range behaviour, this functional is found to grossly underestimate the energy of charge-transfer states. They suggest that it is better to use the CIS method with scaling factor to correct for its systematic overestimation of excitation energies. The implication of the importance of charge transfer states inferred from the aforementioned results of Jean and Hall should therefore be considered with caution.

The Matsika Group studied excited state dynamics in 2AP|Y[22] and 2AP|R[23] nucleobase dimers. A second-order Møller-Plesset perturbation correction was included in calculations to enhance the CIS method (which is known to be deficient in handling electron correlation effects). A typical B-DNA conformation was used as a starting geometry for calculations. Optimisation was also performed on a modified dimer where the 2AP monomer was replaced with its first excited state equivalent. Both 5'-2AP|N-3' and 5'-N|2AP-3' dimer permutations were considered. Switching the base order resulted in differences in the overlap of the bases as well as the specific interbase interactions that were possible. The fate of the excited state was determined by following relaxation pathways across the S_1 potential energy surface and meant that the dominant fluorescence quenching mechanism could be exposed for each system. Exciplex formation, conical intersections, charge transfer, and dark states were all implicated as possible relaxation mechanisms available to the excited state. Importantly, the pathway that was favoured was dependent on not only the identity of the base neighbouring 2AP but also the initial conformation of the excited state;

specifically, different behaviour was observed for dimers that had the 2AP monomer in the 5' position compared to those where it was in the 3' position.

The studies by the Matsika Group did not account for the effect of the sugar-phosphate backbone or solvation and so may be difficult to relate to experimental observations. Additionally, geometric changes from the initial B-DNA conformation were attributed to relaxation in the excited state; however, as the global minimum for a ground state dimer structure is not of B-DNA form, these observations are perhaps of limited value. Nevertheless, the key observation is that different relaxation mechanisms were available to the dimer depending on the conformation of the excited state. This is of great significance to studies involving systems where conformational fluctuations are known to occur. The complex nature of the fluorescence decay of 2AP incorporated into a nucleic acid construct could be caused by excited state dynamics as well as sampling of various ground state conformations.

5.1.2 Structures Investigated in this Study

A key objective of this investigation was to interrogate the stability of 2AP-containing dinucleotides which were in a conformation similar to the canonical structure of a DNA duplex. The presence (or absence) of a stable, stacked conformation would help to elucidate whether or not the short lifetime component observed in a 2AP dinucleotide could be due to a similar quenching mechanism found in a duplex. Stacking parameters were obtained from DNA crystal structures to provide standard values to which comparison of computational results could be made.

It was necessary to establish a full and detailed description of the structures as there were many possible conformational degrees of freedom. For instance, the bases in a dinucleotide could be stacked closely together but their relative orientation might have been significantly distorted from the canonical base-stacking geometry. Such perturbed conformations might be a possible source for intermediate fluorescence lifetimes. Either charge transfer would be less efficient (giving a smaller rate constant) or some molecular reorganisation would have to take place to allow the dinucleotide to adopt a conformation where charge transfer was possible. Seemingly subtle differences in the conformation of a dinucleotide compared to a duplex may therefore have significant photophysical consequences.

As previously outlined, fluorescence lifetime measurements of 2AP-dinucleotides suggest that there is a small population of 2AP which is in an environment similar to that experienced by free 2AP in solution. Therefore another important aspect of this

investigation was to determine whether or not it was possible for a dinucleotide to exist in a stable, open conformation (that is, a geometry that is analogous to a flipped base in a DNA duplex) which could account for this observation.

It is worth highlighting that, unlike many previous studies, the dinucleotide sugar-phosphate backbone was included in calculations performed in this investigation. This greatly increased the size and complexity of the system; however, it also meant that the bases were constrained in a more biologically relevant way. Some of the consequences of including the backbone were investigated by considering the specific interactions in which it was involved. Particular attention was paid to the phosphate group which naturally carries a negative charge in solution.

2AP dimer structures (that is, base-stacked structures without a covalent link between the two bases), designated 2AP|N, were investigated to provide insight into the stacking interaction between the bases when configured in a dinucleotide-optimised conformation. Unrestricting the nucleobases during the optimisation meant that the effect of the sugar-phosphate backbone in dinucleotides could be inferred. Skeleton dimers were also investigated to determine how generic interactions influenced the conformation of the bases. This helped to uncover how much of an effect the exocyclic functional groups had on the stacking geometry.

As mentioned previously, 2AP has many desirable qualities as a probe of DNA conformation and dynamics; however, as it is not a natural base, there has always been a concern that its incorporation into a duplex might cause significant perturbation to the native structure. Computational calculations provided an opportunity to investigate subtle changes that might occur in the structure of a dinucleotide (and, by extension, a DNA duplex) when adenine is replaced by 2AP. This was achieved by comparing each 2AP-N dinucleotide to its analogous adenine dinucleotide (A-N). A similar comparison was also made between 2AP|N and A|N dimer structures.

The effect of solvation was considered by comparing structures optimised in water and gas-phase environments. Calculations were also performed in solvents other than water but these results will not be discussed in detail.

5.2 Methods

5.2.1 Molecular Structure

All dinucleotides studied were of the form 5'-2AP-N-3' or 5'-A-N-3' (where N = adenine (A), guanine (G), cytosine (C), or thymine (T); the natural DNA bases). Preliminary geometries for 2AP-N dinucleotides, which were provided by collaborator Tanja van Mourik,[24] were prepared by mutating adenine to 2AP in appropriate dinucleotides extracted from a DNA crystal structure in the Protein Data Bank (PDB). An initial optimisation in water (see below, §5.2.2) provided the starting geometry for any subsequent calculations performed on the structures (for example, when using a different PCM solvents). It should be noted that the dinucleotides contained only a single phosphate group, the one bridging the two nucleotides. The terminal phosphate was excluded to be consistent with experimental samples. A schematic of a typical dinucleotide structure used during this investigation is shown in Figure 5.1.

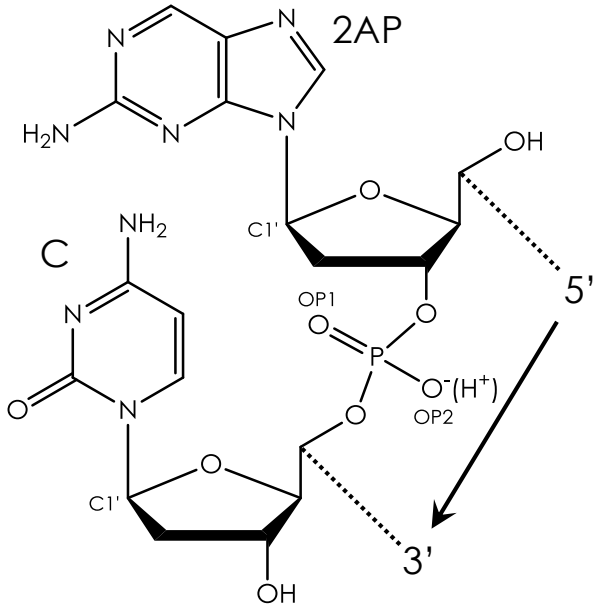

Figure 5.1 – 5'-2AP-N-3' dinucleotide structure used in this study. In the dinucleotide shown, N is cytidine, C. For reference, the C1' carbon atoms of the deoxyribose sugar groups are labelled. Note that there is only one phosphate group in the molecule. Calculations were performed on structures with and without a proton cap on the phosphate group. Structures with OP2 capped are denoted PO4H while structures with OP1 capped are denoted altPO4H.

The starting geometry of each A-N dinucleotide was created by mutating 2AP back to adenine in the analogous, optimised 2AP-N structure. Of course, the original

dinucleotide structure cut from DNA could have been used for this purpose but it was predicted that the 2AP-N geometry would be closer to the optimal conformation and, thus, would converge faster.

In the case of 2AP-T a number of different stacked geometries were investigated. Early calculations provided two distinct conformational families; one where there was an intramolecular H-bond (HB2) in the backbone of the dinucleotide (3R86_HB2) and one where this H-bond was not present (3R86). A further 2AP-T family was obtained when a different DNA crystal structure was used to create the preliminary geometry (4C64). The notation 4C64 and 3R86 refers to the PDB codes for the crystal structures from which the dinucleotides were derived. The H-bond HB2 will be discussed in more detail in a later section. The difference between the stacked 2AP-T families is illustrated in Figure 5.2.

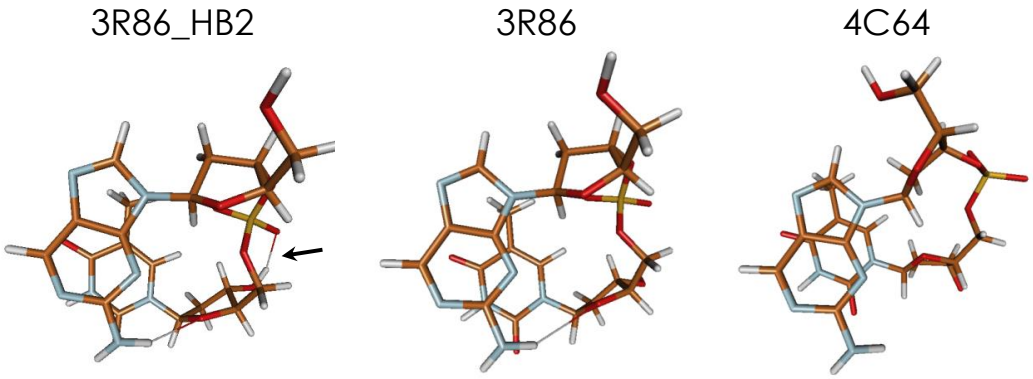

Figure 5.2 – Comparison of 2AP-T conformational families 3R86_HB2, 3R86, and 4C64. The black arrow highlights the position of the backbone H-bond (HB2) that differentiates the 3R86_HB2 and 3R86 conformations. The 4C64 structure exhibits significantly different base-stacking to the other two conformations.

An open (flipped[f]) state was estimated for 2AP-T using the crystal structure of a complex of DNA and methyltransferase *Thermus aquaticus* I (M. TaqI).[25] The DNA duplex contained the appropriate recognition sequence (TCGA, where the underlined adenine is the target base for methylation) for base-flipping with the exception that the adenine had been replaced with 2AP. The M. TaqI crystal structures also provided a 2AP-G dinucleotide where the 2AP had base-paired in a Hoogsteen geometry.[26,27]

[f] The use of 'flipping' and 'flipped' in a structure containing only two bases may seem inappropriate; however, these terms will still be used for discussion in the context of duplex structure.

An initial attempt to determine the conformational (base-flipping) pathway between stacked and open 2AP-T structure led to yet another family of 2AP-T structures. In this case the bases were perpendicularly oriented (T-shaped). While the base-flipping transition was not of major concern to this investigation it was concluded there was some merit in comparing the results of the T-shaped structure with the more traditional stacked structures.

The negatively charged phosphate groups in the backbone of DNA are generally balanced with counter-ions (for example Na^+, K^+, Mg^{2+}, or Ca^{2+}) in solution. Modelling a cation was potentially problematic as, if left free, it would be able to diffuse around the system and could disrupt calculations. Fixing the cation might also have influenced calculations as a decision would have to be made about where it should be placed. If the conformation of the dinucleotide changed significantly then the position of the fixed cation might become inappropriate. Additionally, simulations using molecular dynamics suggest that direct binding of counter-ions is actually quite sporadic and so fixing the cation would perhaps be an unsuitable protocol for this system.[28] It was therefore concluded that it would be better to leave the single phosphate group charged or cap one of its two non-bridging oxygen atoms with a proton.

Nucleic acid dimer starting structures were created from optimised dinucleotide geometries. Except for the C1' carbon atoms, the phosphate-sugar backbone was completely removed. The C1' atoms were subsequently mutated into hydrogen atoms. Further removal of the functional groups from these nucleobase dimers also provided purine (Pu) and pyrimidine (Py) skeleton dimer structures. These skeleton structures only contained ring atoms and exocyclic hydrogen atoms. All dimer structures had neutral charge as they did not contain phosphate groups.

5.2.2 Computational Calculations

Geometry optimisations were performed in Gaussian 09[29] with functional M06-2X[30] and basis set 6-31+G(d). The geometry optimisations utilised Gaussian's 'tight' convergence criterion and 'ultrafine' integration grid (containing 99 radial shells and 590 angular points per shell). Harmonic vibrational frequencies were computed at the same level of theory to verify the nature of the stationary points (minima or transition states) and to compute zero-point energy corrected interaction energies (E_0). Solvent effects were included using the polarizable continuum model (PCM). Solvents studied (with their dielectric coefficient, ε) were as follows; water (H_2O, ε = 78.3553), dimethylsulfoxide (DMSO, ε = 46.8260), ethanol (EtOH, ε = 24.8520) and

dichloromethane (DCM, ε = 8.9300). Calculations were also performed without PCM solvent to simulate gas-phase (*in vacuo*) measurements.

Structures were first optimised using H_2O as the PCM solvent. This optimised structure then acted as a starting geometry for all further calculations. The only exception to this was the 3R86_HB2 2AP-T structures. In this case the starting geometry was obtained from the optimisation of the 3R86 form when using DCM as the PCM solvent.

5.2.3 Analysis of Stacking Geometry

Molecules were visualised and manipulated using a combination of Jmol,[31,32] PyMol,[33] and Molden.[34,35] Base-stacking geometry was investigated using 3DNA,[36,37] w3DNA,[38] or an in-house script written in Fortran (supplied by collaborators[39]). This script was also ported to MATLAB[40] for ease of use.

The programs 3DNA and w3DNA (which will herein be collectively known as x3DNA programs) are based on methods from earlier work[41–43] and conform to standard definitions for nucleic acid structure.[44] Schematic diagrams of the base axis definitions used are shown in Figure 5.3 and Figure 5.4. Dinucleotides (and dimers) were analysed as single-stranded DNA using the definitions of Strand I. A combination of Open Babel[45] and MATLAB was used to produce the Protein Data Bank format (pdb) input files required for the programs.

For all structures the errors in base-stacking parameter values were estimated as the following: shift, slide, and rise ≤ 0.5 Å, tilt and roll ≤ 1°, and twist ≤ 2°. It is worth mentioning that these estimates are based on the accuracy of the computational calculations rather than the x3DNA analysis programs. The accuracy of a calculation can be influenced by many factors including the choice of functional, basis set, and convergence criteria.

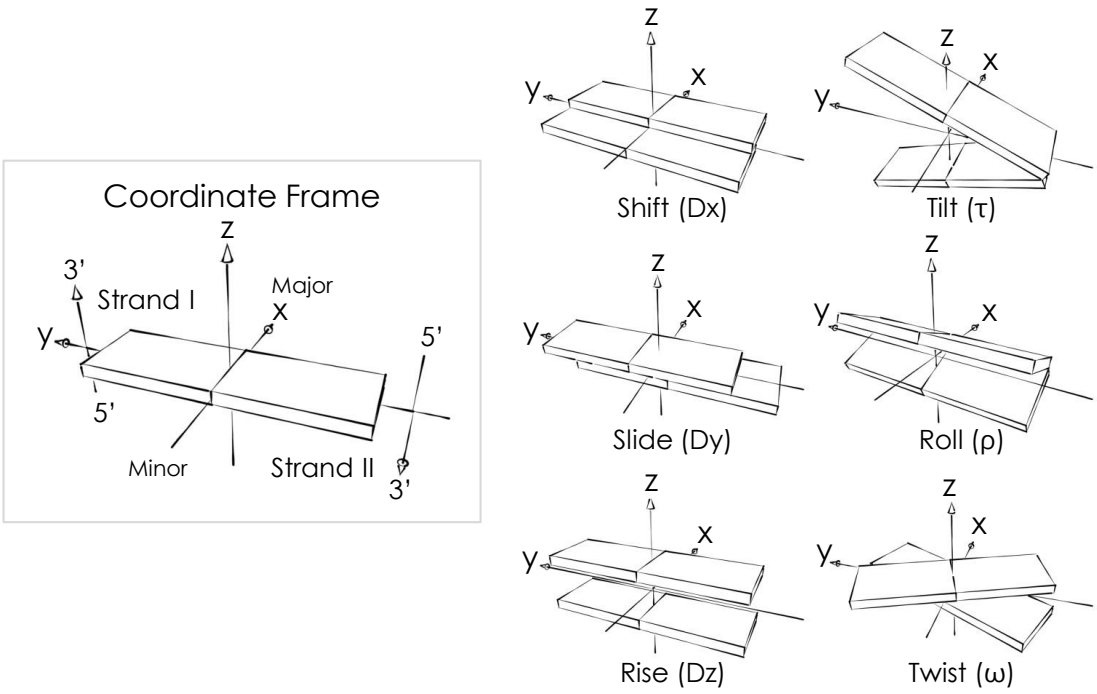

Figure 5.3 – Parameter definitions used by analysis x3DNA programs. The coordinate frame shows definitions relative to minor and major DNA grooves. The models were based on those of Lu and Olson[37] and created using SketchUp.[46]

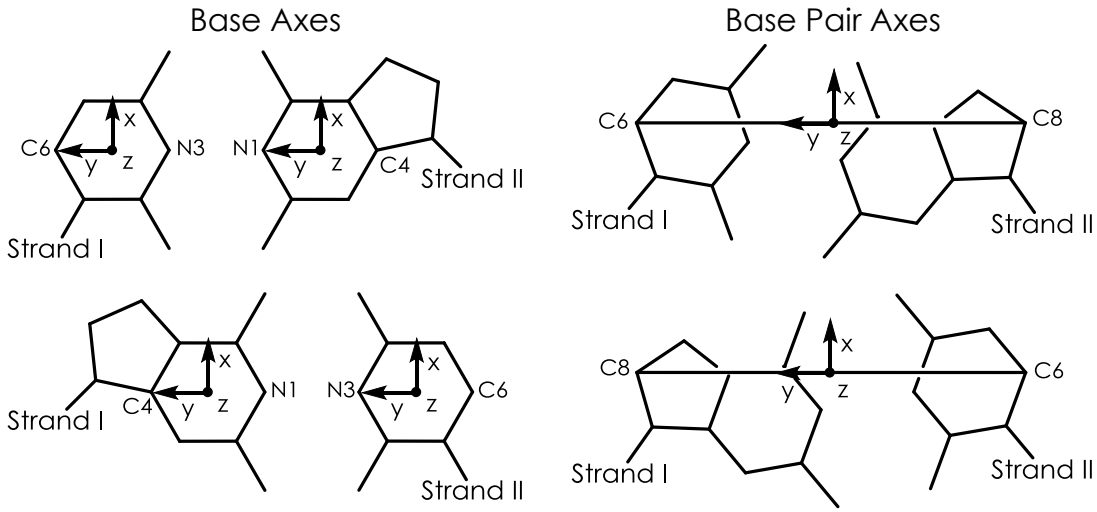

Figure 5.4 – Axis definitions used by x3DNA programs for purine and pyrimidine bases. A full explanation of the scheme is given by El Hassan and Calladine.[42] In brief, the y-axis points along the direction of N1-C4 (purine) or N3-C6 (pyrimidine) from Strand II to Strand I. The centre of the line between the two y-axis defining atoms gives the origin of the axes. The normal of the plane of the base (pointing away from the page as shown) defines the z-axis. The x-axis is defined by completing a right-hand set. Similar definitions, which should be clear from the diagrams, give the base pair axes.

5.3 Results and Discussion

This section will have a particular focus on putting the results in the context of DNA duplex structure. Some consequences of the specific system used for calculations will also be explored. For example, the effect of capping or not capping the phosphate group will be considered. The differences between structures optimised in water and the gas-phase will also be investigated. Following this there will be a brief discussion of the more unusual dinucleotide conformations observed (flipped 2AP-T, T-shaped 2AP-T, and Hoogsteen 2AP-G).

5.3.1 Anionic Dinucleotide or Capped Phosphate Group?

A summary of the parameters obtained for anionic dinucleotides (No Cap) and those that had a hydrogen atom (proton) capping one of the two available oxygen atoms on the phosphate group (PO4H or altPO4H, see Figure 5.1) is presented in Table 5.1. Slide, shift, rise, tilt, roll, and twist parameters are those defined in Figure 5.3. ΔE_0 is the energy difference between capped structures and μ_{Total} is the total dipole moment of the system.

Dinucleotide	Shift /Å	Slide /Å	Rise /Å	Tilt /°	Roll /°	Twist /°	ΔE_0 /kJ mol^{-1}	μ_{Total} /Debye
2AP-A								
No Cap	1.5	−0.1	3.1	0	0	61	N/A	8.0
PO4H	1.6	−0.1	3.2	0	0	61	0.0	13.3
altPO4H	1.6	−0.2	3.2	−1	0	60	−3.3	9.5
2AP-C								
No Cap	0.5	−1.5	3.7	−9	−1	46	N/A	9.2
PO4H	1.4	−0.1	3.3	−2	−4	57	0.0	15.4
altPO4H	0.2	−1.5	3.5	−4	−4	43	−4.2	12.6
2AP-G								
No Cap	1.5	−0.2	3.1	5	−4	59	N/A	14.0
PO4H	1.6	−0.1	3.1	4	−4	60	0.0	10.2
altPO4H	1.6	−0.2	3.1	4	−4	60	−2.4	6.9
2AP-T								
No Cap	−0.1	−1.6	3.4	−2	−6	39	N/A	7.3
PO4H	−0.1	−1.1	3.3	0	−6	40	0.0	17.0
altPO4H	−0.1	−1.6	3.4	−1	−6	39	−6.4	12.3

Table 5.1 – A comparison of base-stacking parameters obtained for anionic (No Cap) or capped (PO4H or altPO4H) dinucleotides in water.

The similarity of the values suggests that having an overall negative charge had little effect on the optimised structure. In fact, changing the capping position from one phosphate group oxygen atom to the other caused about the same magnitude of change in the structure.

Based on Mulliken charges the phosphate group of the anionic dinucleotides was found to have an overall charge of approximately −1 (the phosphorous atom held a charge of +2.1 to +2.3 while the surrounding oxygen atoms had a combined charge of between −3.1 and −3.3). This suggests that the negative charge remained localised to the phosphate group during the optimisation. This was desirable because delocalised charge might have interfered with intramolecular interactions. When capped, the phosphate group had an overall charge of around −0.1 to −0.2 (the phosphorous atom had a charge of +2.2 to +2.3, the hydrogen cap held a charge of around +0.6 and the surrounding oxygen atoms had a combined charge of between −3.0 and −3.1).

One value that did change quite significantly between anionic and capped structures was the dipole moment. Interestingly (with the exception of 2AP-G) the dipole moment increased when the phosphate group was capped. Figures 5.5-5.8 show the optimised structures of Table 5.1 and, in addition, relevant dimer structures (2AP|N), which are discussed in more detail in §5.3.4. The electrostatic potential has been mapped onto the solvation shell for each structure. Colour scale shows negative values as red and positive values as blue. The black arrow represents the total dipole moment of the system.

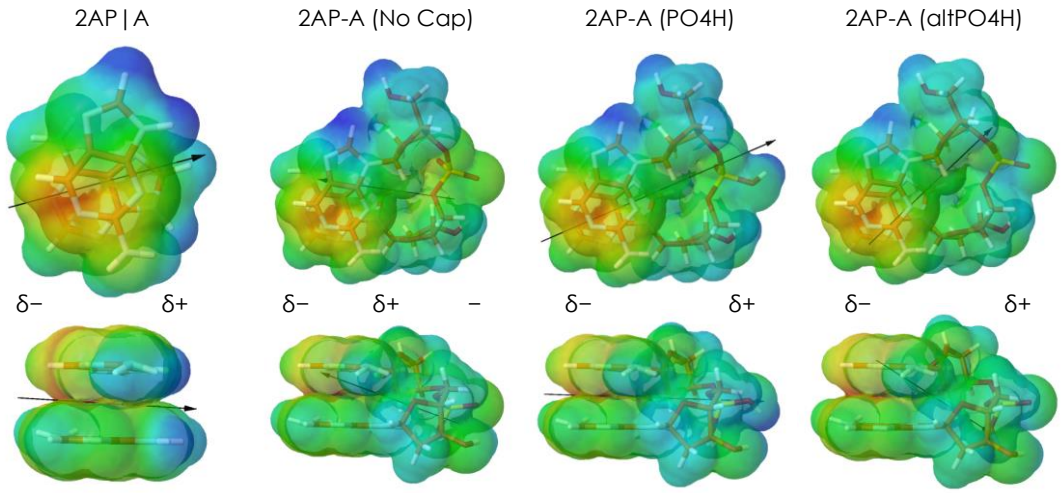

Figure 5.5 – Optimised adenine-containing structures with the electrostatic potential mapped onto the solvation shell. A crude representation of the charge distribution is given by the symbols between the two perspectives of the structures. In the No Cap structure the phosphate group contains a full negative charge and appears to neutralise the natural charge heterogeneity of the 2AP|A dimer. Capping the phosphate group causes an increase in the magnitude of the total dipole moment. Despite the stark changes observed in the dipole moment base-stacking is essentially the same in all structures.

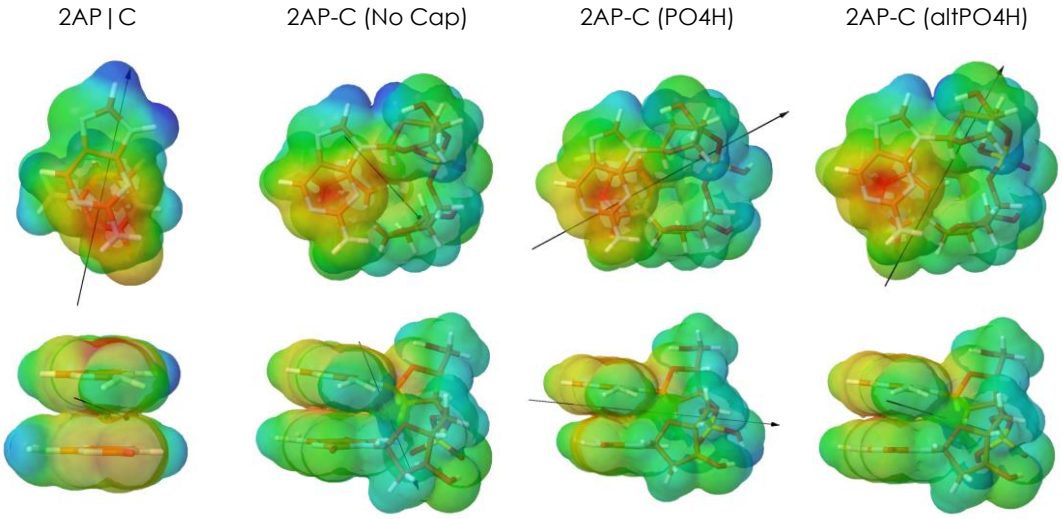

Figure 5.6 – Optimised cytosine-containing structures with the electrostatic potential mapped onto the solvation shell.

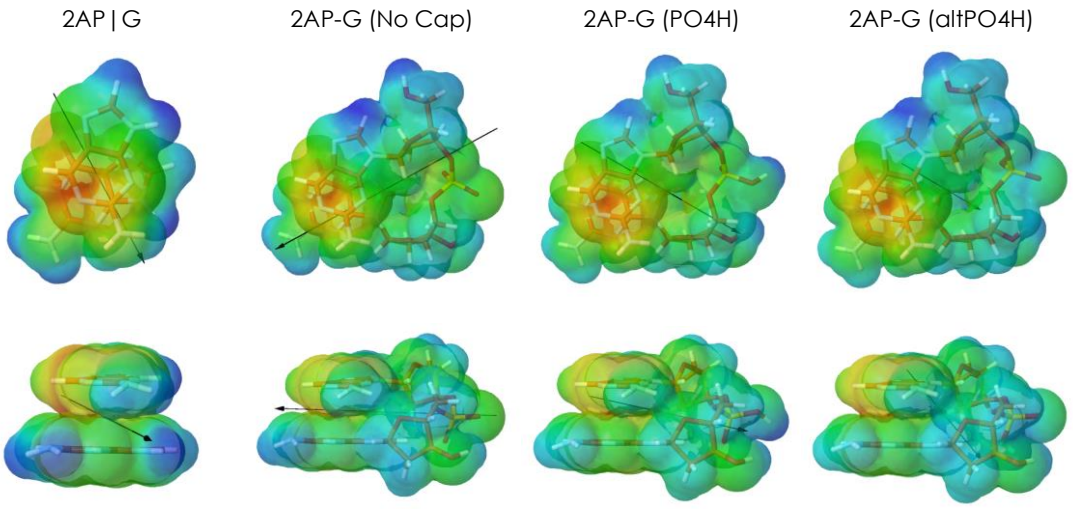

Figure 5.7 – Optimised guanine-containing structures with the electrostatic potential mapped onto the solvation shell.

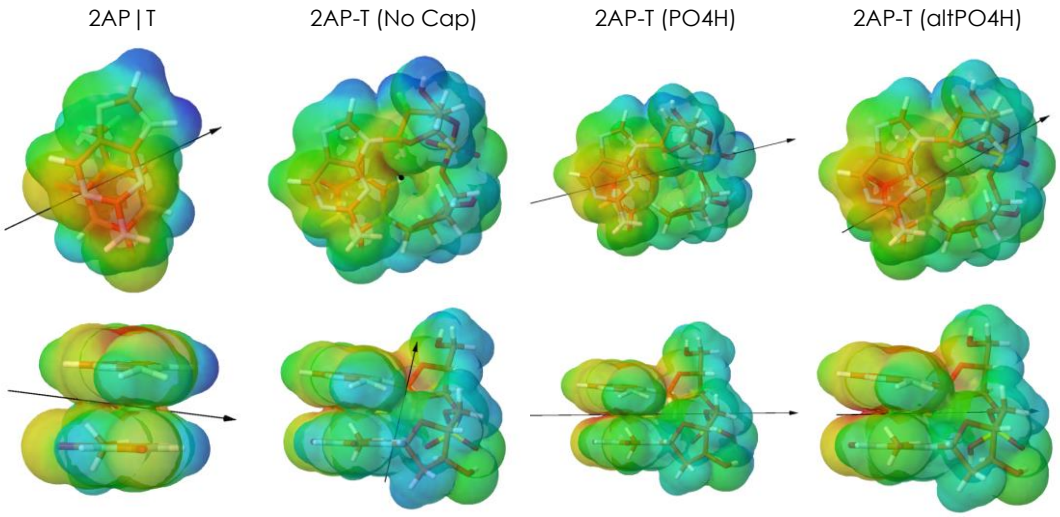

Figure 5.8 – Optimised thymine-containing structures with the electrostatic potential mapped onto the solvation shell.

In all cases there is a dramatic change in dipole direction between dimer and anionic dinucleotide structures. When the phosphate group is capped (on either oxygen) the dipole of the dinucleotide reverts to a direction similar to that of the dimer. For adenine, cytosine, and thymine structures the addition of the charged phosphate group appears to counter the charge heterogeneity of the dimer. Thus, when the charge is neutralised by capping the phosphate group the overall dipole moment increases. In the case of guanine the charged phosphate group enhances the charge heterogeneity. Capping the guanine dinucleotide therefore causes a reduction in the overall dipole moment.

Capping the phosphate group clearly has a large impact on the overall dipole moment of the dinucleotide but base-stacking (and, in fact, the general geometry of the structure) is essentially unchanged. Given that there was little difference between anionic and capped structures, it was concluded that it would be uneconomical to repeat calculations for the same dinucleotide with and without the phosphate cap. Therefore, unless otherwise stated, the results all pertain to uncapped, anionic dinucleotides.

5.3.2 Base-Stacked Structures of 2AP-N Dinucleotides in Water

For visualisation the structures are presented as ball and stick models. Colours for atoms are as follows; carbon, brown; oxygen, red; nitrogen, blue; phosphorous, orange; and hydrogen, white. H-bonds are represented as thin connectors between the appropriate atoms. It should be noted that the H-bonds are interpreted by the visualisation package and their presence (or absence) may be misleading (H-bond visualisation parameters were as follows: distance, 1.50-3.15 Å; angle, 145-215°). It should also be noted that the structures are shown with 2AP above the other base (N); however, in accordance with convention, the rise is calculated from 2AP to N (downwards as shown). It would be advisable to interrogate structures using a visualisation package if more detail is sought. Optimised structures in Cartesian format (XYZ) can be found in Supplementary Information (CD:\ComputationalStructures\).

H-bonds that were commonly observed in 2AP-N dinucleotide structures are highlighted in Figure 5.9. HB1 is an inter-nucleotide H-bond between the 2AP amino group (H22) and the neighbouring deoxyribose sugar group (O4'). HB2 is an intra-backbone H-bond between the terminal H3' atom and the O2P oxygen atom of the phosphate group. It should be noted that H3' would not be present in a duplex and so an analogous H-bond would not be possible.

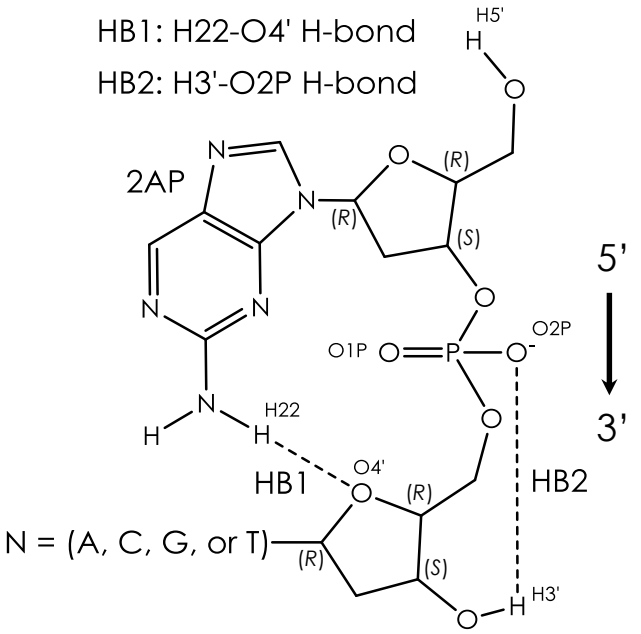

Figure 5.9 – Schematic highlighting some internal H-bonding common to many of the stacked 2AP-N dinucleotides. The stereochemistry of chiral carbon atoms is given by R/S notation.

Due the similarity between the 2AP-R (R = A or G) structures, respectively, and the 2AP-Y (Y = C or T) structures, respectively, the two types of dinucleotide will be discussed separately. A qualitative description of the geometries will be followed by a more quantitative analysis of stacking parameters.

5.3.2.1 2AP-R Structures

Figure 5.10 and Figure 5.11 show the geometry-optimised structures obtained for 2AP-A and 2AP-G, respectively. In both cases the two bases of the dinucleotide are almost parallel to each other and are significantly overlapped. An H-bond appears to be present between the 2AP amino group and the neighbouring purine sugar group (HB1).

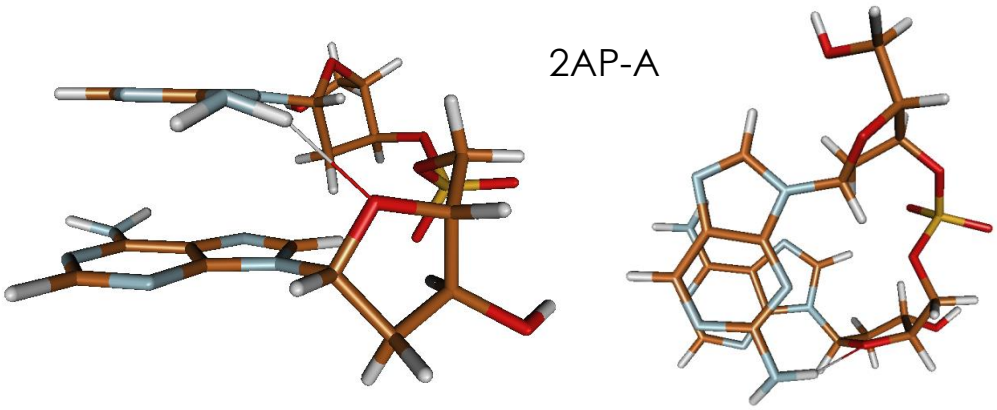

Figure 5.10 – Optimised geometry for 2AP-A in water. Left: Side view. Right: Top view.

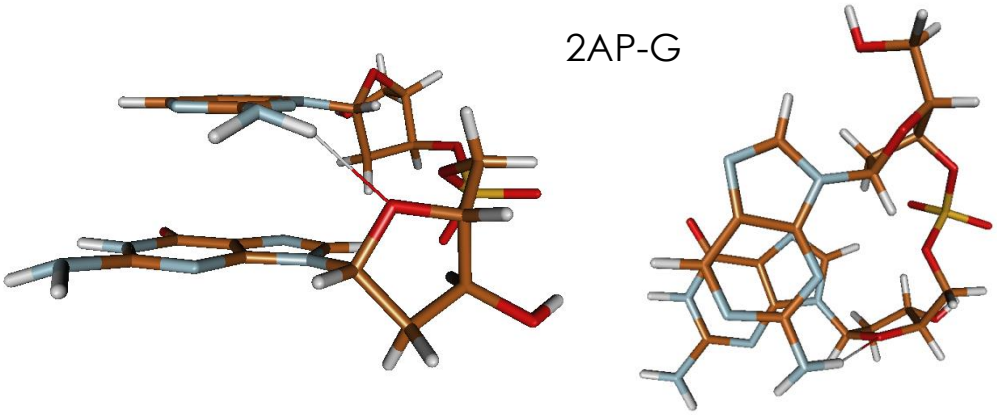

Figure 5.11 – Optimised geometry for 2AP-G in water. Left: Side view. Right: Top view.

5.3.2.2 2AP-Y Structures

Figure 5.12 and Figure 5.13 show the geometry-optimised structure obtained for 2AP-C and 2AP-T, respectively. Similar to the 2AP-R structures, the two bases of the dinucleotide are significantly overlapped. There is, however, a much more obvious tilt between the bases – that is, they are not as close to being parallel to each other. Again, there is an apparent H-bond between the 2AP amino group and the sugar group of the neighbouring base (HB1). In addition, 2AP-Y structures show an H-bond within the backbone structure. Specifically, there is a bond between the outer phosphate group oxygen (O2P) and the pyrimidine sugar hydroxyl (HB2).

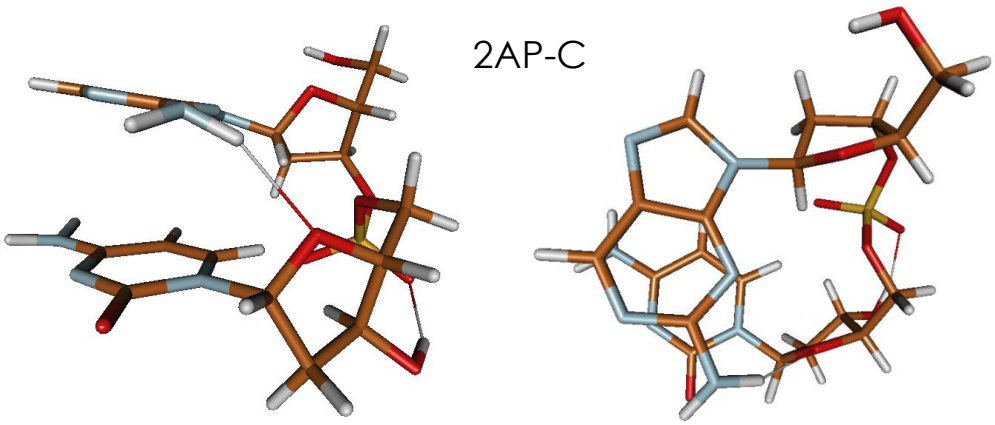

Figure 5.12 – Optimised geometry for 2AP-C in water. Left: Side view. Right: Top view.

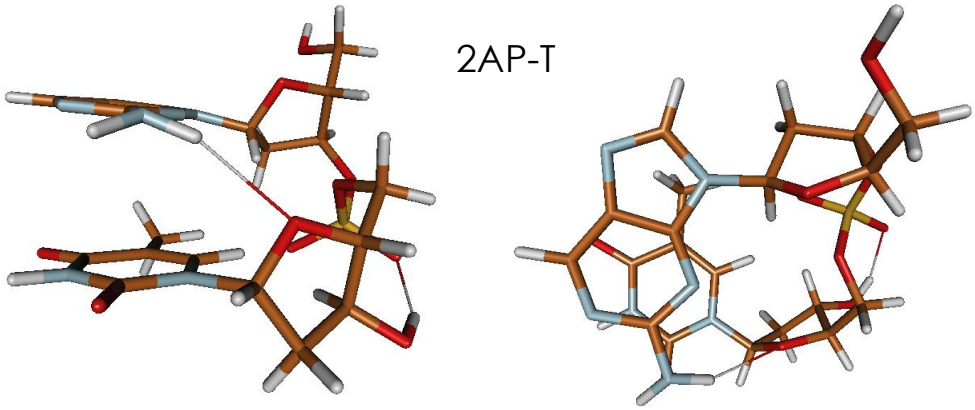

Figure 5.13 – Optimised geometry for 2AP-T in water. Left: Side view. Right: Top view.

5.3.2.3 Comparison of 2AP-N Base-Stacking to DNA Duplex Structure

The stacking parameters of the 2AP-N dinucleotides are given in Table 5.2. Mean values of base-step parameters for (B-form) DNA are also shown; these values were derived by Olson *et al.* from high resolution DNA crystal structures (2.0 Å or better resolution without chemical modification, base mismatches, drugs, or proteins) from the Nucleic Acid Database.[47] It is obviously important to be aware that the structures analysed to obtain the DNA parameters were crystal structures. As such, the DNA structure might have been perturbed from its solution-phase geometry. This could perhaps have resulted in exaggerated distortions in the base-stacking geometry. Using mean DNA values may also be slightly misleading as averaging can obscure conformational heterogeneity. Indeed, idealised values would suggest that the bases should stack directly on top of each other with a 3.34 Å rise and 36° twist; however,

this is not consistent with the fact that (aromatic) stacking usually occurs with some lateral shift, as observed with benzene.[48] Nevertheless, mean DNA values are useful as a point of reference for comparing dinucleotide conformations to what might be expected in larger constructs. It is particularly interesting to note the large standard deviation of the mean twist value (~7°) which suggests that it can vary quite considerably; this is consistent with the idea that certain base-step sequences have a smaller twist angle that is compensated by other base-step sequences.[49] Mean values are also useful because they provide a rough guide of typical values for the stacking parameter definitions used by the x3DNA programs.

Dinucleotide	Shift /Å	Slide /Å	Rise /Å	Tilt /°	Roll /°	Twist /°
DNA						
Idealised[36]	0.00	0.00	3.34	0.00	0.00	36.00
Mean[47]	−0.02	0.23	3.32	−0.10	0.60	36.00
Standard Deviation[47]	0.45	0.81	0.19	2.50	5.20	6.80
2AP-R						
2AP-A	1.5	−0.1	3.1	0	0	61
2AP-G	1.5	−0.2	3.1	5	−4	59
2AP-Y						
2AP-C	0.5	−1.5	3.7	−9	−1	46
2AP-T	−0.1	−1.6	3.4	−2	−6	39

Table 5.2 – Summary of stacking parameters for 2AP-N dinucleotides with H_2O as the PCM solvent. DNA stacking parameters derived from X-ray crystal structures, as well as idealised values, are also provided for comparison; see main text for details.

The values in Table 5.2 corroborate the split between 2AP-R and 2AP-Y geometries evident from the visual inspection of the dinucleotides. 2AP-R dinucleotides have quite small tilt and roll values indicating that the bases are almost parallel to each other. The rise values indicate that base-stacking is more compact than that found in DNA. 2AP-Y stacking is more distorted and the bases are further apart. This distortion is most likely caused by the formation of HB2 in the sugar-phosphate backbone.

Interestingly the twist value of 2AP-A, 2AP-C, and 2AP-G dinucleotides (especially the 2AP-R structures) is very large. A possible explanation of this is that the formation of HB1 (the 2AP-amino to sugar H-bond) pulls the 2AP base around, increasing the twist angle. This effect is discussed in more detail in the next section when a comparison is made to adenine dinucleotides which cannot form HB1 (§5.3.3).

The presence of the phosphate group charge might also be a factor in causing hyper-rotation of the 2AP base. The dipole moment of isolated 2AP (see Chapter 2, §2.3.1, Figure 2.7) suggests that it would be electrostatically favourable for the base to obtain this orientation; however, as capping the phosphate group had little effect on the overall structure (including the twist angle), it would seem that this electrostatic interaction is unlikely to be a major contributing factor. Another potential reason for the apparently large twist is that there is only one stacking interaction. In a duplex the bases are constrained by two neighbours and also by base-pairing. It is therefore not very surprising that the optimal dinucleotide geometry differs to some extent from that of a duplex.

Unfortunately, due to the increased computational cost it would have incurred, it was not possible to investigate extended structures to test if a more constrained system did give more duplex-like stacking parameters. Although larger structures could not be investigated, it was possible to gain some insight into the stacking behaviour of the dinucleotides by looking at the analogous dimer structures. This approach is discussed in a later section (see §5.3.4).

The differences between the base-stacking of the dinucleotide structures obtained in this study and that of canonical duplex conformation might help to explain discrepancies in the fluorescence lifetime behaviour of 2AP in these systems. There would therefore be value in performing photophysical calculations on 2AP base-stacked in each of these geometries to gain insight into this issue.

5.3.3 Comparison of 2AP and Adenine Dinucleotides

Figure 5.14 shows an overlay of the structure of each 2AP-N dinucleotide with that of its analogous adenine dinucleotide (A-N).

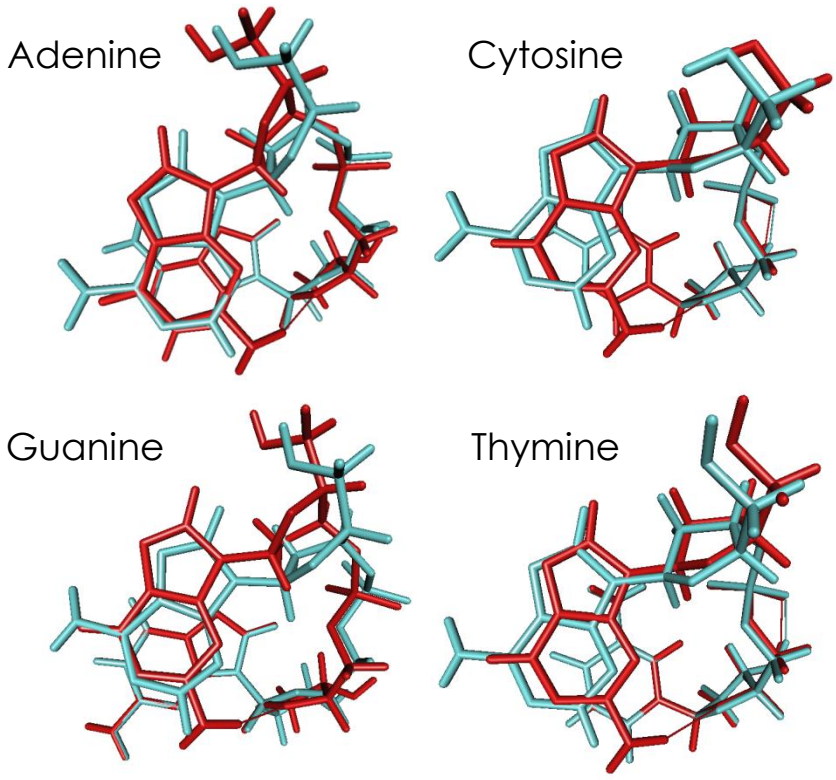

Figure 5.14 – Overlay of the structures of 2AP-N (red) and A-N (blue) dinucleotides. Structures were aligned using the natural 3'-base.

Visual inspection suggests that 2AP can mimic adenine base-stacking well. There is little difference in the conformation of the structures beyond the presence of HB1 in 2AP dinucleotides. This H-bond is not possible in adenine dinucleotides due to the position of the NH_2 group. A comparison of the stacking parameters obtained for these structures is given in Table 5.3.

Dinucleotide	Shift /Å	Slide /Å	Rise /Å	Tilt /°	Roll /°	Twist /°	ΔE_0 /kJ mol^{-1}
Adenine							
2AP-A	1.5	−0.1	3.1	0	0	61	0.0
A-A	1.2	−0.1	3.2	−2	0	52	−9.4
Cytosine							
2AP-C	0.5	−1.5	3.7	−9	−1	46	0.0
A-C	−0.1	−1.2	3.7	−9	7	40	−4.0
Guanine							
2AP-G	1.5	−0.2	3.1	5	−4	59	0.0
A-G	1.5	−0.1	3.2	1	−4	50	−10.3
Thymine							
2AP-T	−0.1	−1.6	3.4	−2	−6	39	0.0
A-T	−0.7	−1.2	3.2	1	−3	34	−3.6

Table 5.3 – A comparison of 2AP-N and A-N dinucleotide structures. ΔE_0 is the energy difference between analogous dinucleotides.

In general, the values are fairly consistent between the 2AP and adenine dinucleotides suggesting that 2AP stacks in a comparable manner to the natural bases. Given the structural similarity of 2AP to both adenine and guanine this was perhaps to be expected. The twist values of both adenine-dinucleotides and 2AP-dinucleotides are quite large compared to DNA stacking (again, particularly the purine samples); however, in all cases adenine dinucleotides have smaller twist values. This difference may be due to the fact that adenine dinucleotides cannot form an H-bond equivalent to HB1 observed in 2AP dinucleotides. It would seem logical that formation of this H-bond might cause a slight relative rotation of the 2AP base compared to adenine. Although, as adenine dinucleotides also exhibit large twist angles relative to DNA, the formation of HB1 is unlikely be solely responsible for hyper-rotation of the 2AP base. It is worth mentioning that guanine-containing dinucleotide structures (of the form 5'-G-N-3') investigated by Barone *et al.* exhibited an H-bond analogous to HB1;[50] this shows that such an interaction is not limited to 2AP-containing dinucleotides and can occur in structures containing only natural bases.

Adenine dinucleotides are lower in energy than equivalent 2AP dinucleotides by an amount that ranges from 3.6 to 10.3 kJ mol^{-1}; this is fairly consistent with the 11.0 kJ mol^{-1} difference observed between free adenine and 2AP nucleobases at the same level of theory, M06-2X/6-31+G(d)/PCM.

5.3.4 Structures of 2AP | N Dimers in Water

Figure 5.15 shows optimised geometries for 2AP-containing nucleobase dimers.

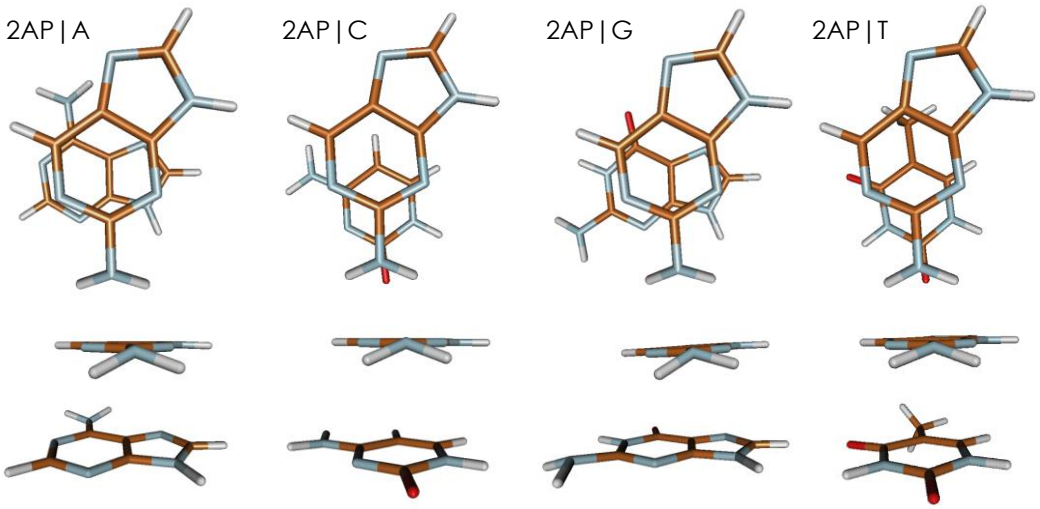

Figure 5.15 – Optimised structures of 2AP|N dimers in water. The top row shows the structures from 'above' while the bottom row shows a 'side' view of the same structures.

In 2AP|R (R = A or G) structures the nitrogen atoms of the 6-membered rings are staggered while in 2AP|Y (Y = C or T) structures they are almost eclipsed (disregarding the relative shift and slide of the bases). It appears that the orientation of the bases in 2AP|Y dimers is influenced by an interaction between the 2AP amino group and the pyrimidine base O2 oxygen atom. This results in a relative rotation of bases such that amino and carbonyl groups are eclipsed and is likely to be an electrostatic effect.

A comparison of stacking parameters for dinucleotides and their equivalent dimer structures is given in Table 5.4.

2AP-N or 2AP|N	Shift /Å	Slide /Å	Rise /Å	Tilt /°	Roll /°	Twist /°	μ_{Total} /Debye
Adenine							
Dinucleotide	1.5	−0.1	3.1	0	0	61	8.0
Dimer	1.6	−0.2	3.1	0	2	58	7.9
Cytosine							
Dinucleotide	0.5	−1.5	3.7	−9	−1	46	9.2
Dimer	−0.9	−2.3	3.1	2	−2	17	11.1
Guanine							
Dinucleotide	1.5	−0.2	3.1	5	−4	59	14.0
Dimer	1.6	−0.2	3.1	4	−4	60	9.0
Thymine							
Dinucleotide	−0.1	−1.6	3.4	−2	−6	39	7.3
Dimer	−0.9	−2.5	3.2	1	−3	16	10.1

Table 5.4 – A comparison of 2AP-N dinucleotide and 2AP|N dimer stacking parameters.

These results again show that there is a clear difference between 2AP|R and 2AP|Y structures. 2AP|R dimer stacking is essentially unchanged relative to the dinucleotide structures. In contrast, the 2AP|Y dimer parameters are significantly altered. 2AP|C and 2AP|T appear to stack more effectively (closer and more parallel) when the backbone is absent. These results suggest that the 2AP-R dinucleotide bases are stacked in an almost optimal position while 2AP-Y dinucleotide base-stacking is significantly distorted. As mentioned previously, this difference is likely due to the formation of HB2 (the backbone H-bond) in 2AP-Y structures. Specifically, distortion of the sugar-phosphate backbone results in sub-optimal stacking.

Twist values are of particular interest in this dinucleotide-dimer comparison. The values obtained for 2AP-R dinucleotides and 2AP|R dimers are almost identical. This suggests that the large twist observed for 2AP-R structures is predominantly due to the optimisation of the stacking interaction of the bases. The 2AP|Y dimer twist values are much smaller than their parent 2AP-Y dinucleotide values. Optimisation of twist angle between bases therefore seems to be constrained by the dinucleotide backbone in 2AP-Y samples. Visualisation of overlaid dinucleotide and dimer structures appears to back up this conclusion. The 2AP|Y dimer structures shown in Figure 5.16 would require very contorted backbones to form equivalent dinucleotides with the bases locked in their dimer orientation.

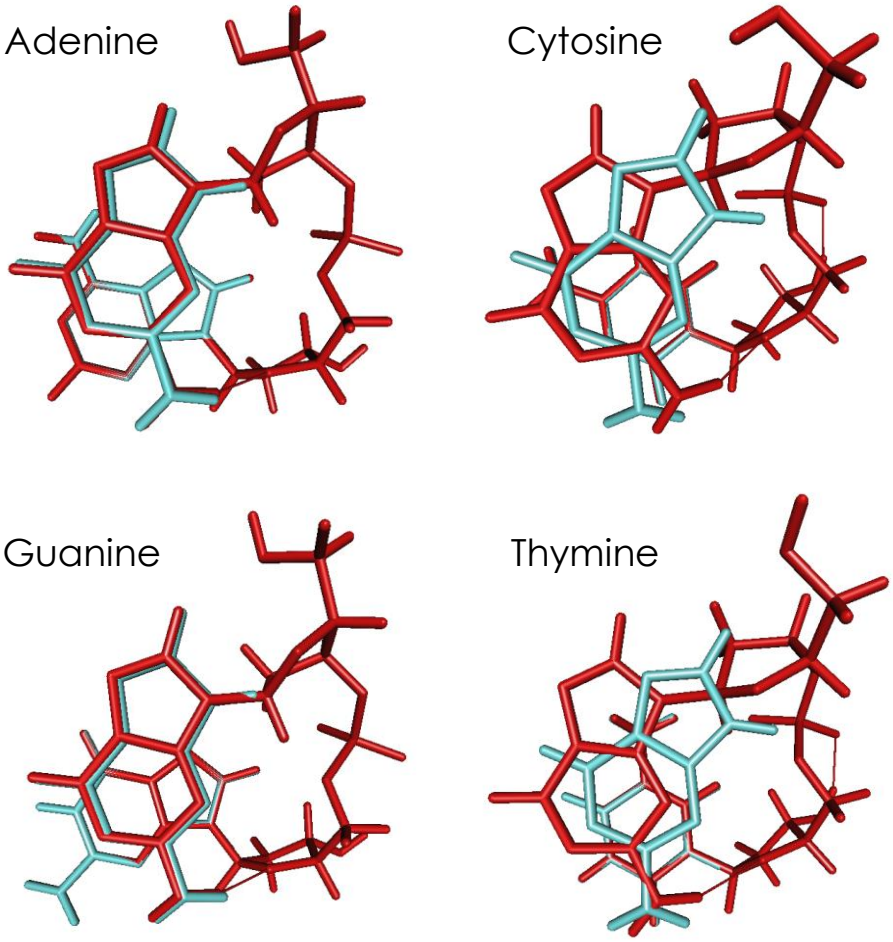

Figure 5.16 – A comparison of stacking in each 2AP-N dinucleotide (red) and its corresponding 2AP|N dimer (blue). Structures were aligned to the natural base.

5.3.4.1 Purine and Pyrimidine Skeleton Dimers

Figure 5.17 shows optimised geometries for purine and pyrimidine skeleton dimers.

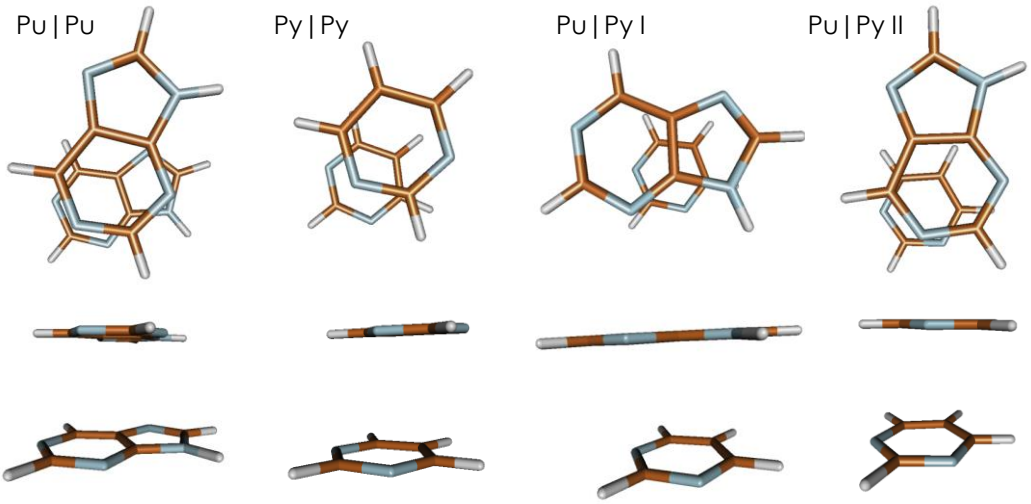

Figure 5.17 – Optimised purine (Pu) and pyrimidine (Py) skeleton dimer structures. Two geometries are shown for the Pu|Py heterodimer. The top row shows the structures from 'above' while the bottom row shows a 'side' view of the same structures.

Purine (Pu|Pu) and pyrimidine (Py|Py) homodimers show similar nitrogen staggering to that observed in 2AP|R structures. Two conformations were found for the purine|pyrimidine (Pu|Py) heterodimer when different initial geometries were used.

A summary of stacking parameters for the skeleton dimers is given in Table 5.5.

Structure	Shift /Å	Slide /Å	Rise /Å	Tilt /°	Roll /°	Twist /°	μ_{Total} /Debye
Skeleton Dimers							
Pu|Pu	1.4	−0.2	3.2	−1	−1	58	8.9
Py|Py	3.1	−0.8	3.1	3	−4	67	5.4
Pu|Py I	−0.8	−0.6	3.2	0	−3	3	6.2
Pu|Py II	2.7	−1.8	3.3	−1	0	83	8.2

Table 5.5 – A comparison of stacking of purine (Pu) and pyrimidine (Py) skeleton dimer structures. Two Pu|Py heterodimer geometries were analysed.

The parameters obtained for the skeleton dimer structures are generally consistent with those of the nucleobase dimers. Interestingly both of the heterodimer structures are perturbed from the nitrogen staggered geometry favoured by the homodimers. While Pu|Py II appears to visually show nitrogen staggering in Figure 5.17 the twist value calculated indicates a significant distortion from the expected value of around 60°. This suggests that, in addition to the effect of the exocyclic functional groups, there is perhaps an inherent difference between stacking in 2AP|R and 2AP|Y structures.

5.3.4.2 Comparison of 2AP|N and A|N Dimers

Overlaid structures of 2AP- and adenine-containing dimers are shown in Figure 5.18. Associated stacking parameters for these structures are given in Table 5.6.

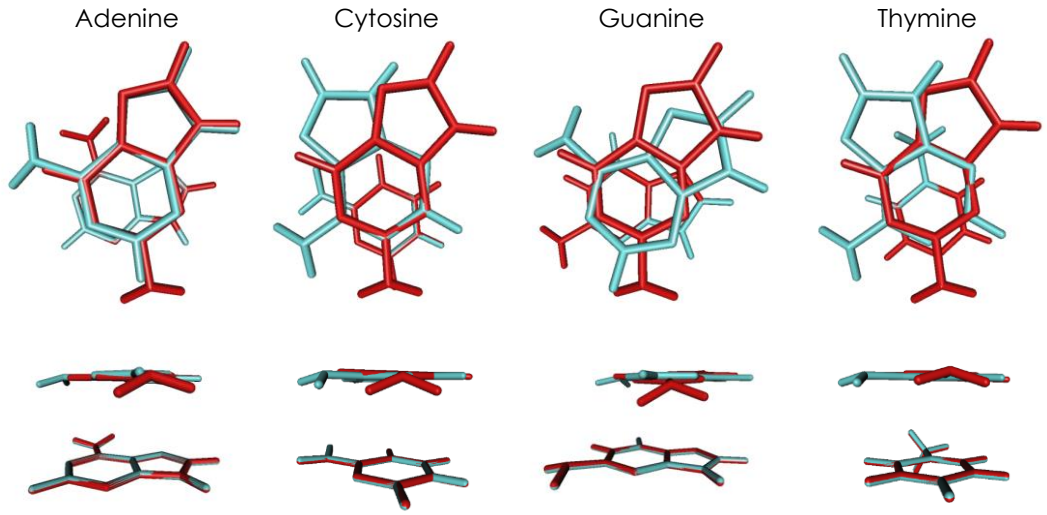

Figure 5.18 – A comparison of 2AP|N (red) and A|N (blue) dimers. Structures were aligned to the natural base. The top row shows the structures from 'above' while the bottom row shows a 'side' view of the same structures.

Dimer	Shift /Å	Slide /Å	Rise /Å	Tilt /°	Roll /°	Twist /°	ΔE_0 /kJ mol^{-1}	μ_{Total} /Debye
Adenine								
2AP|A	1.6	−0.2	3.1	0	2	58	0.0	7.9
A|A	1.4	−0.2	3.1	0	4	56	−8.6	6.1
Cytosine								
2AP|C	−0.9	−2.3	3.1	2	−2	17	0.0	11.1
A|C	2.4	−1.9	3.1	1	3	75	−11.3	12.6
Guanine								
2AP|G	1.6	−0.2	3.1	4	−4	60	0.0	9.0
A|G	1.4	−0.3	3.1	2	2	43	−7.3	6.9
Thymine								
2AP|T	−0.9	−2.5	3.2	1	−3	16	0.0	10.1
A|T	2.2	−2.0	2.9	6	−2	69	−12.7	7.2

Table 5.6 – A comparison of structural parameters for 2AP|N and A|N dimers. ΔE_0 is the energy difference relative to the 2AP|N dimer.

The results again suggest that 2AP stacks in a similar way to adenine. The rise values show that base separation is almost identical and in all cases (except perhaps

A|T) bases appear to be close to being parallel to each other (evident from small tilt and roll values).

There is, however, a significant discrepancy when comparing twist values. This is particularly true for pyrimidine structures where there is almost 60° separating optimal twist orientations. The position of the NH_2 group therefore clearly plays an important role in determining the precise stacking geometry of the dimer. The NH_2 group in adenine is in the C6 position and so (at least for the starting geometry used in this case) the O2 oxygen atom of the pyrimidine base is relatively far away. As a consequence the amino-carbonyl interaction which appears to occur in 2AP|Y structures does not appear to be relevant for A|Y dimers. The lack of the amino-carbonyl interaction (and perhaps the inclusion of some other interbase interaction) causes the A|Y dimers to take on a distinct geometry from the 2AP|Y dimers. In fact, the A|Y dimers assume a conformation similar to that of the Pu|Py II skeleton heterodimer geometry discussed in §5.3.4.1. For convenience a summary of stacking parameters for these structures is given in Table 5.7.

Dimer	Shift /Å	Slide /Å	Rise /Å	Tilt /°	Roll /°	Twist /°
A|C	2.4	−1.9	3.1	1	3	75
A|T	2.2	−2.0	2.9	6	−2	69
Pu|Py II	2.7	−1.8	3.3	−1	0	83

Table 5.7 – Comparison of stacking parameters for adenine|pyrimidine dimers and a skeleton purine|pyrimidine dimer.

2AP|R dimers have larger twist values that their analogous A|R dimers. Therefore the larger twist angle in 2AP dinucleotides compared to adenine dinucleotides is perhaps due combination of the preferred intrinsic stacking orientation of the bases as well as formation of HB1.

Similar to the dinucleotide system, the energy of adenine dimers is consistently lower than that obtained for analogous 2AP dimers; this time ranging by an amount from 7.3 to 12.7 kJ mol^{-1}. Again, this is comparable to the energy difference observed between free adenine and 2AP nucleobases (11.0 kJ mol^{-1}).

5.3.5 Structure and Interactions of the Sugar-Phosphate Backbone

It is clear from the preceding section that the backbone influences the optimal dinucleotide geometry. Figure 5.19 shows three optimised 2AP-T conformations which all have stacked bases but are structurally distinct from each other. It should be noted

that, unless otherwise stated, discussion about the 2AP-T dinucleotide elsewhere in this chapter refers to the 3R86_HB2 structure. This conformation was chosen to be the default 2AP-T dinucleotide because of its resemblance to the 2AP-C dinucleotide and the fact it was derived from the same initial crystal structure as the other main dinucleotides studied.

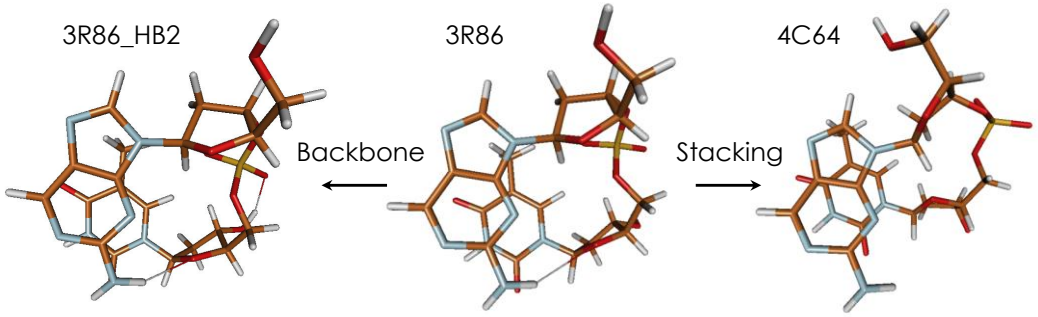

Figure 5.19 – Three distinct (minimum-energy) structures of the 2AP-T dinucleotide. The 3R86 and 3R86_HB2 structures differ primarily in backbone structure while 3R86 and 4C64 structures differ in stacking conformation.

The structures differ mainly in the H-bonds defined in Figure 5.9, namely HB1 and HB2. Stacking parameters and the relative energies of these structures are given in Table 5.8.

Conformation	Shift /Å	Slide /Å	Rise /Å	Tilt /°	Roll /°	Twist /°	ΔE_0 /kJ mol^{-1}	μ_{Total} /Debye
2AP-T								
3R86_HB2	−0.1	−1.6	3.4	−2	−6	39	−11.0	7.3
3R86	−0.3	−0.9	3.3	0	−7	37	0.0	5.6
4C64	0.8	−0.4	2.9	5	0	30	−14.5	11.5

Table 5.8 – Comparison of stacking parameters of 2AP-T dinucleotide structures. The difference in energy, ΔE_0, is given relative to the 3R86 sample.

There is almost 15 kJ mol^{-1} difference in energy between the most and least stable structures. Comparison between 3R86_HB2 and 3R86 structures suggests that formation of HB2 offers a significant stabilisation effect (11 kJ mol^{-1}). This can be considered an optimisation of the backbone structure. The 4C64 structure has neither HB1 nor HB2 but, presumably due to a more favourable stacking interaction, has the lowest energy of the structures found when H_2O was used as the PCM solvent. It is worth pointing out that the 2AP|T dimer structure optimised from the base-stacked geometry of the 4C64 dinucleotide conformation was found to be 5.4 kJ mol^{-1} more

stable than the equivalent 3R86_HB2 derived dimer. This supports the idea that base-stacking in the 4C64 dinucleotide is more stable than that of the 3R86_HB2 structure.

The dinucleotide structures can therefore be stabilised by creating favourable interactions with (or within) the backbone or by improving base-stacking. The availability of (at least) two stabilisation schemes means that the final structure is dependent on the initial structure and also the specific environment in which the calculation is performed (for example, which PCM solvent is used).

Another consequence of these competing stabilisation pathways is that many energy minima could potentially be found for each structure. This means that the optimised structures presented here are not necessarily global minima. While this is important to bear in mind it does not significantly affect conclusions made in this study. It would, however, be very useful to identify the global minima structures in order to infer which conformation(s) would be most populated under experimental conditions.

5.3.5.1 Sugar Pucker and Glycosidic Bond Angle

The specific structure of the sugar-backbone phosphate was not a major focus of this investigation but there are a few points which are worth bringing to light.

As with base-stacking parameters, the sugar pucker (Figure 5.20) and the glycosidic bond angle (Figure 5.21) of DNA are often described in homogeneous terms. The sugar pucker is typically found in a C2'-endo[g] conformation and the glycosidic bond angle is often simply stated to be *anti* (where $|\chi| > 90°$).

[g] Although *exo*/*endo* notation has been superseded by E/T notation (and pseudorotational analysis) its use still appears to be prevalent. It is used here as it is the default output of the x3DNA programs.

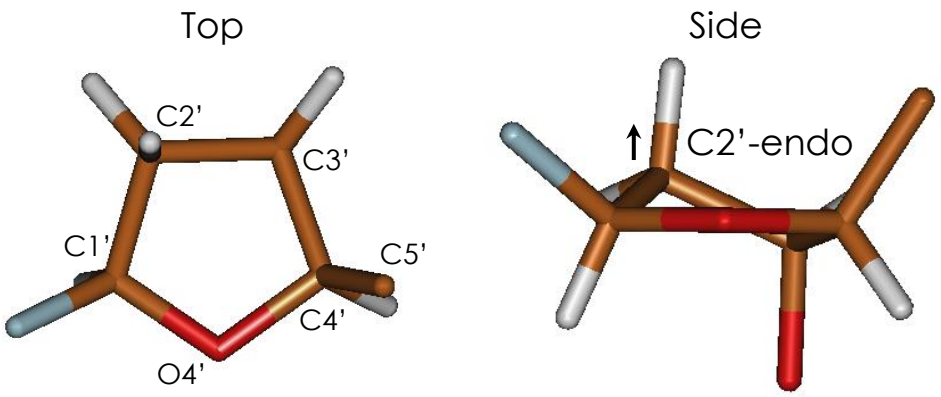

Figure 5.20 – The most common deoxyribose sugar pucker conformation found in DNA is C2'-endo. In this form the C2' atom is found above (as shown) the plane of the sugar group.

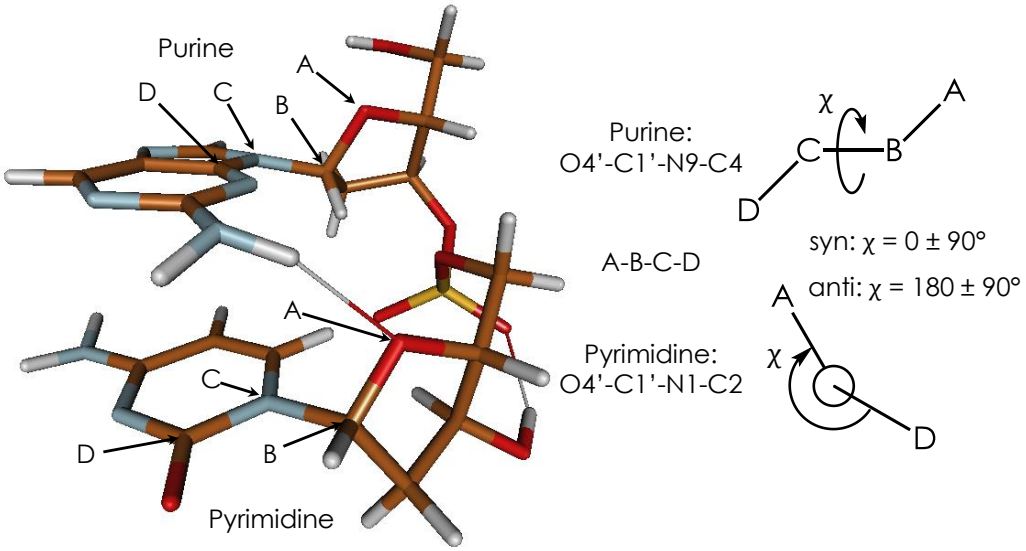

Figure 5.21 – Definition of the glycosidic bond angle (χ). DNA almost always has a χ value which is anti; |χ| > 90°.

The backbone structure of a DNA duplex is often more complicated than this idealised description though. The sugar pucker and glycosidic bond angles generally cover a (limited) range of values.[44] This conformational diversity was realised in the structures calculated in this investigation. Various different sugar pucker conformations were observed; namely, C1'-exo, C2'-endo, C3'-endo, and C3'-exo which are typical of DNA and C2'-exo (gas-phase only), O4'-endo, and O4'-exo which are anomalous for DNA. It should be noted that the O4'-endo and O4'-exo conformations

were also observed in the DNA duplex structures from PDB crystal structures 4C64 and 2M2C. While the average glycosidic bond angle was about −110° there was a fairly large range (around −80° to −140°) sampled by the structures. In general the dinucleotide backbone structure was found to be similar to that observed in a duplex.

Importantly, these observations again highlight the vast complexity of nucleic acid structure. It would appear that the idea of DNA as a specific, well-defined structure is perhaps too simple. Such a description might be sufficient in many circumstances but, when considering the dynamic nature of DNA, it will be important to be aware of the significant variation which is possible. For instance, a recent study of RNA found that a change in sugar pucker conformation was likely necessary to allow enzymatic function to occur.[51]

5.3.6 Effects of PCM Solvent: Structures in Water and the Gas-Phase

5.3.6.1 Comparison of 2AP-N Dinucleotide Structures in Water and the Gas-Phase

Figure 5.22 shows overlays of 2AP-N dinucleotide structures optimised with and without PCM solvent. As a reminder, water optimised structures were used as the starting geometry for gas-phase calculations and so the results only provide insight into differences in a local minimum.

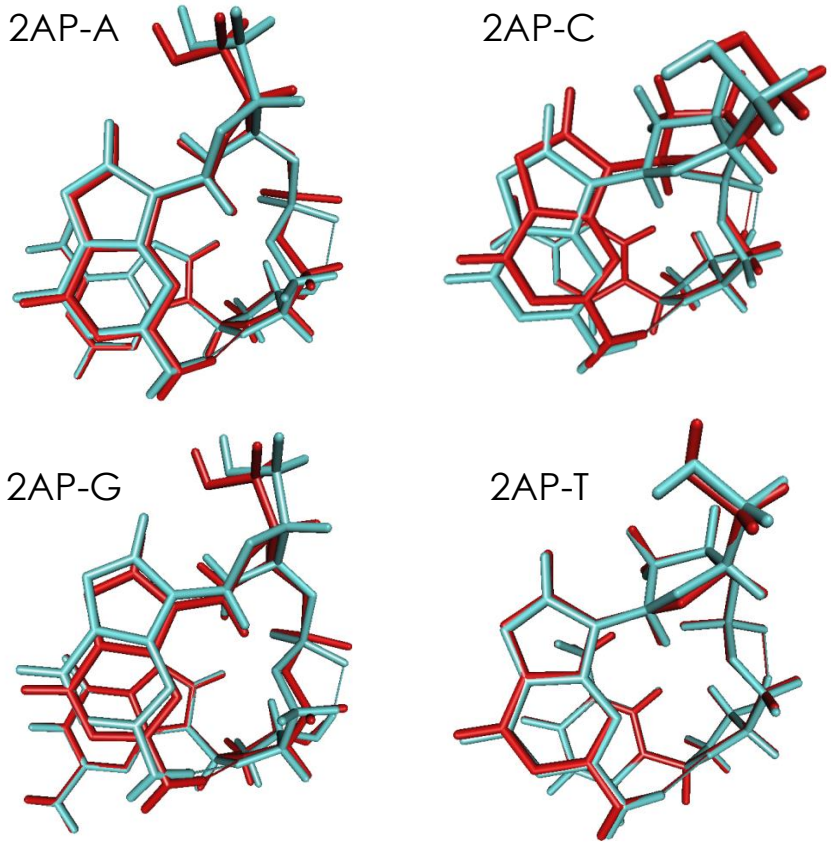

Figure 5.22 – Comparison of optimised dinucleotide structures when the PCM solvent was H_2O (red) or excluded (blue). Structures were aligned using the natural base.

An additional backbone H-bond (HB2) is present in 2AP-A and 2AP-G structures in the gas-phase but, in general, base-stacking is quite similar to that found in water. 2AP-Y structures have HB2 present when PCM solvent is included and so the presence of this H-bond cannot be credited for any changes observed. In the case of 2AP-T there is very little difference in conformation except for rotation of the terminal 5'-hydroxyl group. The 2AP-C dinucleotide shows greater variation between solution- and gas-phase environments but the changes are still fairly minimal.

The base-stacking parameters obtained for the dinucleotide structures in water and the gas-phase are given in Table 5.9.

Dinucleotide	Shift /Å	Slide /Å	Rise /Å	Tilt /°	Roll /°	Twist /°
2AP-A						
H_2O	1.5	−0.1	3.1	0	0	61
Gas-Phase	1.3	−0.1	3.0	4	1	60
2AP-C						
H_2O	0.5	−1.5	3.7	−9	−1	46
Gas-Phase	0.6	−1.0	3.5	−5	0	48
2AP-G						
H_2O	1.5	−0.2	3.1	5	−4	59
Gas-Phase	0.9	−0.4	3.0	10	−5	55
2AP-T						
H_2O	−0.1	−1.6	3.4	−2	−6	39
Gas-Phase	0.1	−1.6	3.2	1	−5	40

Table 5.9 – A comparison of base-stacking parameters in water and the gas-phase.

Although 2AP-R base-stacking conformation is largely maintained in the gas-phase the formation of HB2 causes a slight distortion and greater tilt and roll values are observed. The 2AP-T structure parameters are almost unchanged. Base-stacking in 2AP-C appears to become more dominant in the gas-phase; tilt and roll values decrease implying that the bases are closer to being parallel to each other.

In all of the structures the rise between the bases slightly decreases when solvent is removed. It would seem that in the absence of solvent stabilisation there is an increase in the strength of attractive van der Waals forces between bases.

The relative stability of the different stacked 2AP-T conformations highlighted in the previous section was also influenced by the absence of PCM solvent. 2AP-T structures optimised without PCM solvent are shown in Figure 5.23. The 3R86 starting structure optimised to form HB2 in the gas-phase rather than remaining distinct from the 3R86_HB2 structure.[h] In the case of the 4C64 starting structure, formation of HB1 resulted in a stacking conformation more similar to the 3R86_HB2 structure; however, the orientation of the thymidine sugar group prohibited formation of HB2.

[h] Incidentally, this is how the 3R86_HB2 conformation was initially obtained except that DCM was being used as the PCM solvent. The smaller dielectric coefficient of DCM was enough to favour the formation of the intra-backbone H-bond (HB2).

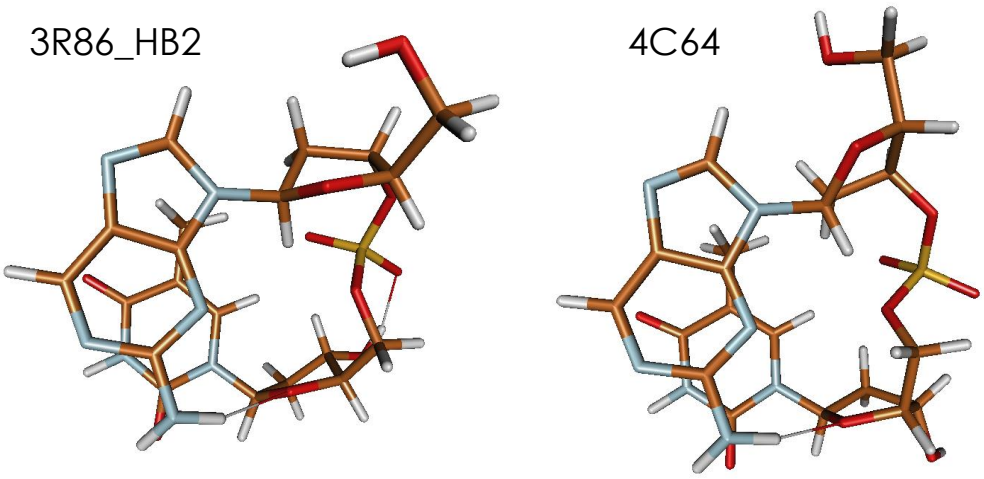

Figure 5.23 – Comparison of stacked 2AP-T dinucleotide structures when PCM solvent was excluded.

Table 5.10 contains a summary of stacking parameters for the distinct 2AP-T dinucleotide conformations found in solution and gas-phase environments.

2AP-T	Shift /Å	Slide /Å	Rise /Å	Tilt /°	Roll /°	Twist /°	ΔE_0 /kJ mol^{-1}
3R86_HB2							
H_2O	−0.1	−1.6	3.4	−2	−6	39	0.0
Gas-Phase	0.1	−1.6	3.2	1	−5	40	0.0
4C64							
H_2O	0.8	−0.4	2.9	5	0	30	−3.5
Gas-Phase	−0.2	−1.6	3.0	7	−4	36	26.6

Table 5.10 – Comparison of stacked 2AP-T dinucleotide structures with and without PCM solvent. The difference in energy, ΔE_0, is given relative to the appropriate 3R86_HB2 structure.

Perhaps the most significant aspect of this comparison is the fact that the relative stability of the structures changes order when PCM solvent is removed. The 4C64 structure is more stable in water while the 3R86_HB2 structure is more stable in the gas-phase. This clearly shows that the specific environment in which calculations are performed can have an influence on the optimal dinucleotide conformation.

5.3.6.2 Comparison of 2AP|N Dimer Structures in Water and the Gas-Phase

Figure 5.24 shows overlays of 2AP|N dimer structures optimised with and without PCM solvent.

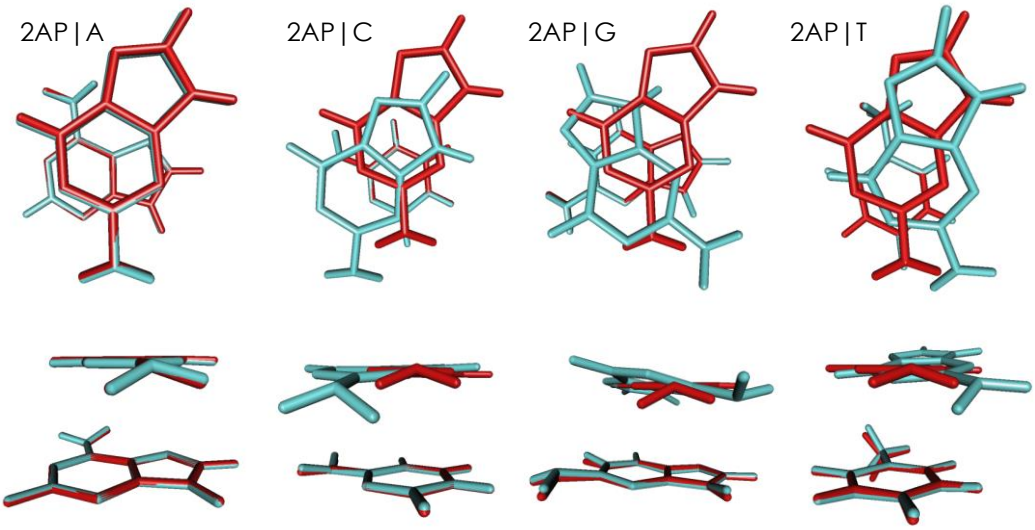

Figure 5.24 – Comparison of optimised 2AP|N dimer structures when the PCM solvent was H_2O (red) or excluded (blue). The top row shows the structures from 'above' while the bottom row shows a 'side' view of the same structures.

In contrast to the dinucleotide structures there is quite a large conformational change caused to 2AP|N dimers when the PCM solvent is excluded. The only exception to this is 2AP|A which shows remarkably similar base-stacking in solution- and gas-phase environments.

The base-stacking parameters obtained for the dimer structures in water and gas-phase are given in Table 5.11.

Dimer	Shift /Å	Slide /Å	Rise /Å	Tilt /°	Roll /°	Twist /°
Adenine						
H_2O	1.6	−0.2	3.1	0	2	58
Gas-Phase	1.6	−0.2	3.1	0	3	59
Cytosine						
H_2O	−0.9	−2.3	3.1	2	−2	17
Gas-Phase	−1.0	−0.7	2.9	6	−6	7
Guanine						
H_2O	1.6	−0.2	3.1	4	−4	60
Gas-Phase	4.2	2.0	3.3	−7	−1	109
Thymine						
H_2O	−0.9	−2.5	3.2	1	−3	16
Gas-Phase	0.9	−2.7	2.6	11	−5	33

Table 5.11 – A comparison of 2AP|N dimer base-stacking parameters in water and the gas-phase.

Consistent with dinucleotide structures, the rise between the bases generally decreases when PCM solvent is removed. Interestingly 2AP|Y dimers appear to adopt particularly small rise values. Visual comparison of the distance between the bases of these structures does not indicate such a significant change is present. This apparent discrepancy may be explained by the greatly increased tilt and roll values of these structures. The 3DNA algorithm uses the centre of the 6-membered ring as a reference point for each base. Thus, changes in the relative inclination of the bases can potentially lead to an effective change in the rise value. This inclination effect could also explain why, in contrast to the other dimers, 2AP|G shows an apparent increase in rise value when the PCM solvent is removed.

Consideration of the electrostatic potential maps of the individual nucleobases (see Chapter 2, §2.3.1, Figure 2.7 and also Appendix 5, §5.1.6) would seem to suggest that the dimer structures re-orientate in the gas-phase to optimise Coulombic interactions. Without solvent stabilisation it is plausible that electrostatic interactions play a more important role. In contrast to the dinucleotide structures there is no backbone to restrict the rearrangement of the bases in the dimers.

5.3.6.3 Dinucleotide Structures with a 5'-Terminal Phosphate in Water and the Gas-Phase

The use of a monophosphate dinucleotide in this chapter is consistent with the structures investigated experimentally in Chapter 3. It is important to realise, however, that this has implications when relating the results to typical DNA structure (or oligonucleotides of three or more units) as the interaction between neighbouring phosphate groups is not included. This means that the structures investigated are somewhat of a special case. In solution the phosphate charge would be (at least partially) neutralised by counter-ions but it was felt that it would be appropriate to determine the effect that including the terminal phosphate would have on the optimisation. Figure 5.25 shows the various structures obtained for monophosphate and diphosphate dinucleotides when the PCM solvent was H_2O (top row) or excluded (bottom row).

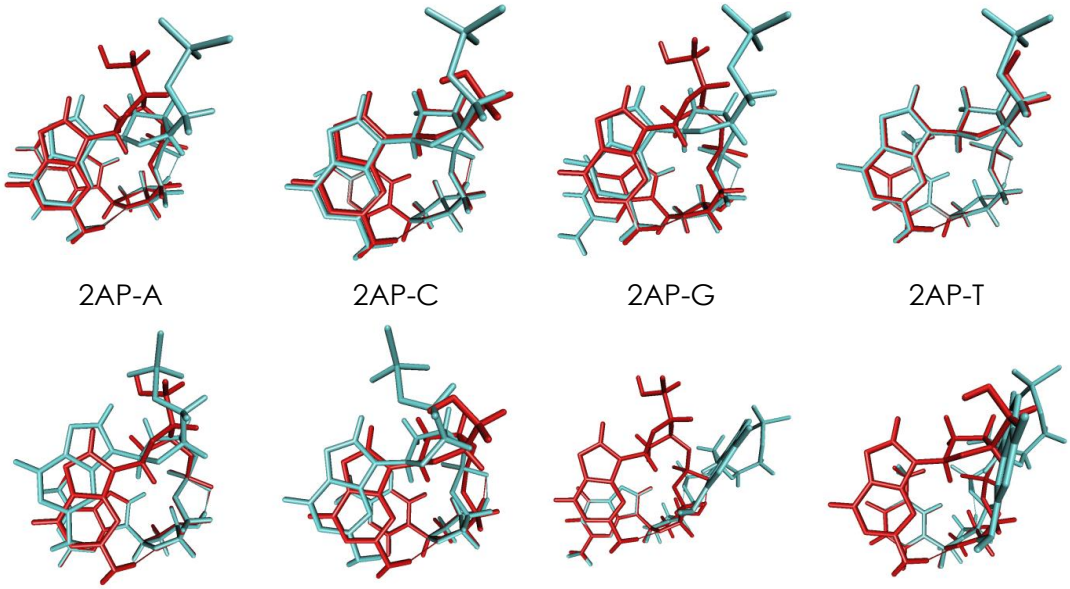

Figure 5.25 – Comparison of optimised dinucleotide structures with one (red) or two (blue) phosphate groups. The top and bottom rows show the geometry obtained when the PCM solvent was H_2O or excluded, respectively. Structures were aligned using the 3'-base.

The results suggest that when solvent is included the optimal geometry is similar for the monophosphate and diphosphate dinucleotides. The major difference in these structures is that the backbone of 2AP-R diphosphate structures is distorted relative to the monophosphate structures due to the formation of an intramolecular backbone H-bond between H3' and the bridging phosphate (HB2). 2AP-Y monophosphate

dinucleotides already have this H-bond so there is very little perturbation caused when including the extra phosphate.

There is, however, a marked difference when there is no solvent to stabilise the charged phosphate groups (in particular the terminal phosphate which carries −2 charge). For all dinucleotides the terminal phosphate group appears to interact strongly with the H8 atom of 2AP. This has the effect of rotating the 2AP base slightly which, in turn, breaks the H-bond between the 2AP amino group and the neighbouring sugar. 2AP-G and 2AP-T dinucleotides are further distorted as the structure opens out, unstacking the two bases.

It is not completely clear why 2AP-A and 2AP-C remain stacked while 2AP-G and 2AP-T open out but an interbase interaction *via* NH_2 (amino or aniline) groups could perhaps contribute to this difference. 2AP, adenine, and cytosine all have NH_2 groups which could potentially provide a stabilisation effect – see Figure 5.26. It is worth bearing in mind that in the gas-phase there will be no solvent shielding and so polar interactions might be significant across relatively long distances. While guanine also has an NH_2 group it is in a similar position to the 2AP amino group and so this could produce an unfavourable interaction. Thymine lacks an NH_2 group and so perhaps the single NH_2 of the 2AP is not sufficient to hold the bases together. This is, of course, very speculative and further investigation would be required to support this proposition. It is also likely that the effect is not confined to the NH_2 groups but rather is a more general electrostatic interaction.

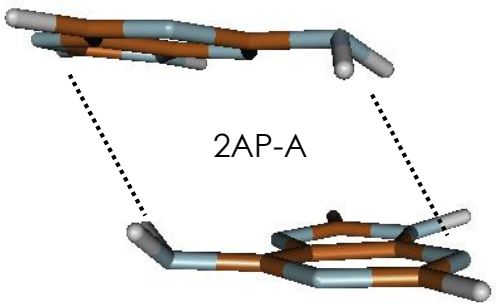

Figure 5.26 – Optimised dinucleotide geometry for 2AP-A when two phosphate groups were present and PCM solvent was excluded. The non-planar geometry of the amino groups suggests that there is some interbase interaction occurring (represented by the dotted line). Note that for clarity the backbone has been removed from the structure.

5.3.7 Open Structures and Base-Flipping

Figure 5.27 shows the geometry optimised structure obtained for flip2AP-T. In this structure the bases are far apart and do not appear to interact with each other or the sugar-phosphate backbone. It therefore seems plausible that a structure of this type could have similar fluorescence character to free 2AP. Visually it is quite difficult to understand why this particular conformation would be an energy minimum; however, one possible explanation is that bond angles and dihedrals are optimised. Further investigation would be required to support this hypothesis but if it was accurate then it would seem likely that there would be a number of similar conformations that would be approximately isoenergetic with the observed geometry.

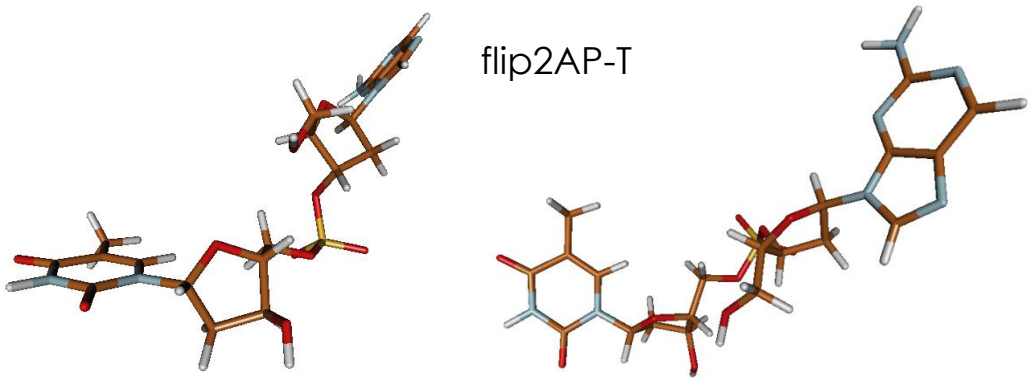

Figure 5.27 – Optimised geometry for flip2AP-T in water. The dinucleotide is in an open structure where the bases do not interact significantly. Left: Side view. Right: Top view.

Figure 5.28 shows a comparison of open 2AP-N dinucleotide structures. All optimised open structures, except for 2AP-T, were supplied by collaborator Leo Holroyd.[52] The similarity between structures is unsurprising given that starting geometries for 2AP-A, 2AP-C, or 2AP-G dinucleotides were obtained by mutating the optimised 2AP-T structure and that the bases are isolated from each other and most of the backbone. Mutation of thymine into one of the other natural bases was therefore unlikely to have much of an influence on the optimal geometry of the structure.

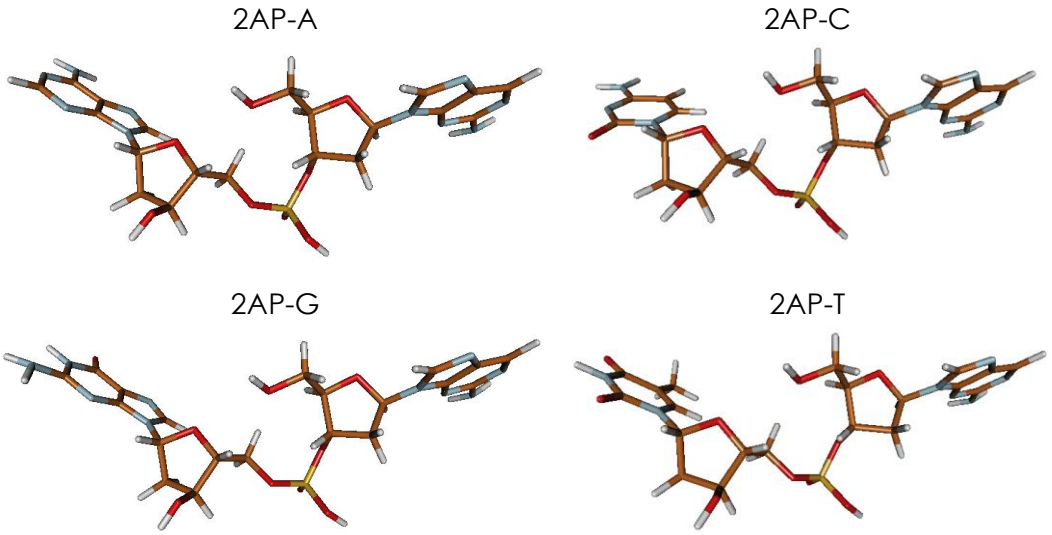

Figure 5.28 – Structures of all open 2AP-containing dinucleotides in the open conformation. The 2AP base is shown on the right in all cases. These structures all have a proton cap on one of the phosphate oxygen atoms. Optimisation of uncapped, anionic structures will be performed in future work.

The energy of the flip2AP-T structure was found to be around 35 kJ mol^{-1} higher than that of the most stable stacked 2AP-T structure. This difference gives a lower-bound to the energy barrier between stacked and open conformations and, as a very rough estimate, seems plausible. The barrier must be significantly larger than thermal energy (~2.5 kJ mol^{-1} at 300 K) as lifetime measurements suggest only a very small population of dinucleotides exist in an open conformation. The energy required for unstacking bases has also been estimated to be at least ~30 kJ mol^{-1} (~7 kcal mol^{-1}) using a number of different experimental techniques including NMR,[53] fluorescence correlation spectroscopy,[54] and fluorescence anisotropy.[55,56] On the other hand, the barrier must still be assailable at room temperature to allow spontaneous base-flipping which is known to occur on the timescale of milliseconds from NMR imino proton exchange experiments.[57–60] It should be noted that the base only has to become accessible to the solvent for imino exchange to occur and so it does not necessarily need to fully flip out of the double helix.[61] As such, kinetic rates for base-flipping are perhaps overestimated from these experiments. Nevertheless, the fact spontaneous base-flipping occurs in a DNA duplex (which is also evident from the phenomenon of DNA breathing[62]) means that the barrier is surmountable at room temperature.

NMR experiments suggest base-flipping in DNA has an activation energy of around 65 kJ mol^{-1} (15 kcal mol^{-1}).[63] Additionally, a number of studies which utilised

molecular dynamics have estimated the free energy difference between stacked and flipped conformations to be around 80 kJ mol^{-1} (20 kcal mol^{-1}).[64–67] In DNA both base-pairing and base-stacking interactions need to be broken and so these estimates can be considered an upper-bound to the energy barrier for a dinucleotide to unstack and open out.

Várnai and Lavery investigated base-flipping in a DNA duplex which contained the recognition sequence for methyltransferase *Haemophilus haemolyticus* I (M. HhaI).[65] Base-flipping is known to be vital for methyl-addition to occur in this system and crystal structures clearly show the target cytosine base flipped out of the double helix and into a catalytic pocket of the enzyme.[68] The molecular dynamics study, which did not include the enzyme in calculations, found that there was no local energy minimum when the base was fully flipped out of the helix. While the stability of the flip2AP-T structure in this study might seem to contradict this outcome, one has to consider the fact that, even in a flipped (or open) state, a nucleotide will have far less freedom in a duplex than a dinucleotide. It is quite possible that a dinucleotide could obtain an open conformation that would be forbidden in a duplex. Thus, while dinucleotides offer a simplified system that can be investigated in more rigorous detail, it might turn out that the pathways between stacked and open conformations are very different to those found in a DNA duplex. This being said, the results of fluorescence lifetime measurements of 2AP-containing DNA and dinucleotides suggest that there is significant similarity between the two systems.

It is clear that a lot more work is required to fully understand base-flipping (opening) in both DNA duplex and dinucleotide constructs. This is going to be a significant challenge due to the complexity (and size) of the system.

5.3.8 Noncanonical Structures

5.3.8.1 Hoogsteen Conformation

During this study some more unusual dinucleotide conformations were also investigated. Figure 5.29 shows the geometry optimised structure for hoog2AP-G. In this structure the 2AP base is rotated around the glycosidic bond (N9-C1') to form a Hoogsteen geometry. In this case, the glycosidic angle is in a *syn* conformation.

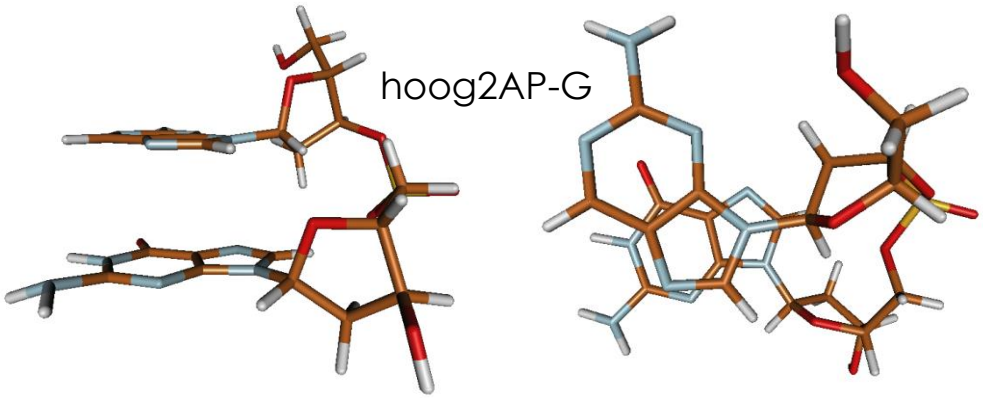

Figure 5.29 – Optimised geometry for hoog2AP-G. Left: Side view. Right: Top view.

Unfortunately, due to the way that the x3DNA algorithm analysed the unusual orientation of the Hoogsteen base, direct comparison could not be made between hoog2AP-G stacking parameters and the Watson-Crick geometry. It was, however, possible to gain some insight into the differences of the conformations by visually comparing them. Figure 5.30 shows an overlay of optimised 2AP-G and hoog2AP-G dinucleotide structures. In contrast to all other stacked conformations, the 5-membered ring (rather than the 6-membered ring) of the (Hoogsteen) 2AP is overlapped with the neighbouring base. Interestingly the 5-membered ring sits exactly where the 6-membered ring would normally be located. There is significant distortion to the backbone structure of hoog2AP-G relative to the Watson-Crick geometry. This is presumably required to obtain the preferred ring position. The amino-sugar H-bond (HB1) observed in other structures is not possible due to the orientation of the 2AP base. The backbone H-bond (HB2) is also prohibited due the orientation of the guanosine sugar group which causes the terminal hydroxyl group to point away from the bridging phosphate.

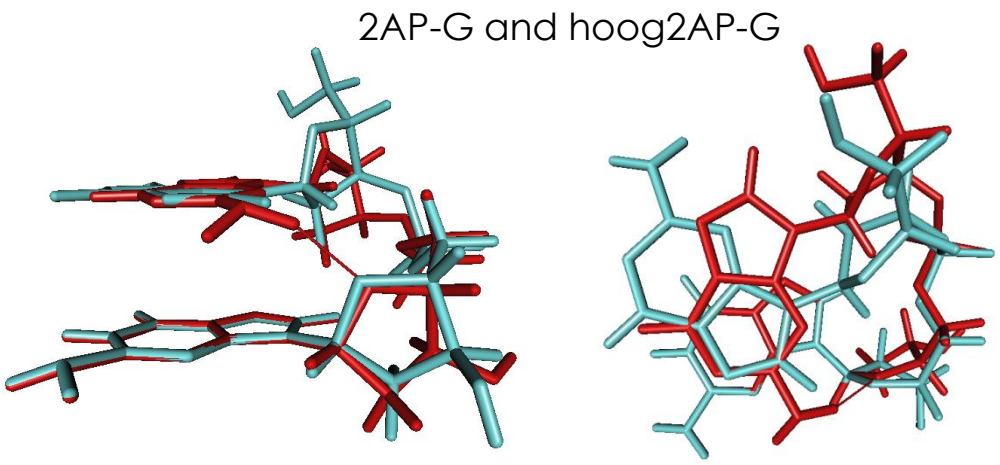

Figure 5.30 – Overlay of optimised 2AP-G (red) and hoog2AP-G (blue) structures. Left: Side view. Right: Top view. Note that the Hoogsteen 2AP 5-membered ring sits inside the Watson-Crick 2AP 6-membered ring. There is a significant difference in the sugar-phosphate backbone of the two structures.

Interestingly the Watson-Crick conformation is only ~5 kJ mol^{-1} more stable than the Hoogsteen geometry. This could potentially mean that a significant proportion of dinucleotides assume the Hoogsteen geometry at equilibrium. Crystal structures also show that 2AP can undertake a Hoogsteen geometry in a DNA duplex.[25] This being said, given the required deformation to the backbone it would seem likely that there would be a much higher energy cost to such a conformational change within a duplex. Its prevalence would also presumably be greatly affected by the specific sequence of nearby bases.

5.3.8.2 T-Shaped Dinucleotide

Another unusual dinucleotide structure was obtained during an initial investigation into the base-flipping (opening) pathway of a dinucleotide. Figure 5.31 shows the geometry optimised structure obtained for 2AP-T in a T-shaped conformation. In this case, the bases are almost perpendicular to each other rather than being in their canonical form. An intra-backbone H-bond (HB2) appears to help to stabilise the observed structure.

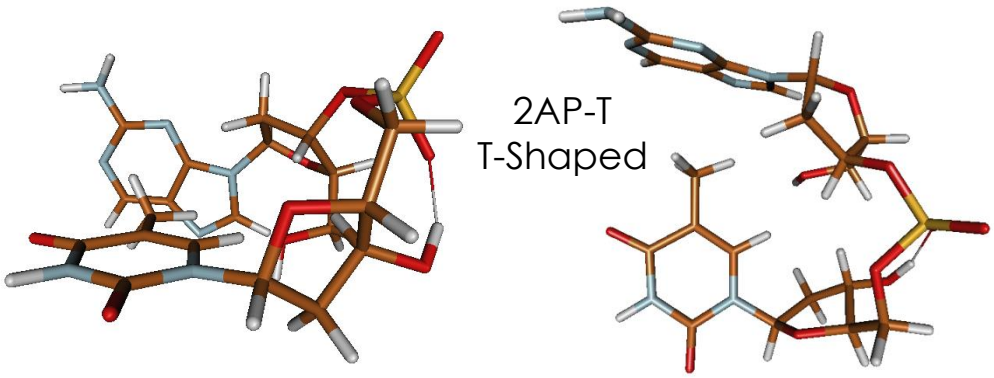

Figure 5.31 – Optimised geometry for 2AP-T in a T-shaped conformation. Left: Side view. Right: Top view.

As previously mentioned base-flipping will be studied in more detail in future studies but there are a few points of interest concerning the T-shaped structure in the context of the current discussion. Table 5.12 gives a comparison of the energies of the various 2AP-T dinucleotide conformations studied.

2AP-T	ΔE_0 /kJ mol^{-1}	μ_{Total} /Debye
Stacked		
3R86_HB2	0.0	7.3
3R86	11.0	5.6
4C64	−3.5	11.5
Distorted		
T-Shaped	2.8	12.5
Open		
flip2AP-T	34.7	10.6

Table 5.12 – Comparison of the zero-point corrected energy and dipole moments of various 2AP-T conformations. Calculations were performed with a PCM solvent of H_2O. ΔE_0 values are all relative to the stacked 3R86_HB2 structure.

It is interesting that the energy of the T-shaped conformation is so close to that of the canonical geometry. One would perhaps intuitively expect that this would be a much less favourable orientation for the bases; however, it has been known for decades that benzene can form stable, T-shaped dimers.[48,69–72] In fact, the energy of a T-shaped benzene dimer is predicted to be close to the global minimum.[73] It therefore seems reasonable that a T-shaped conformation could be a possible energy minimum for a dinucleotide. Obviously this T-shaped dinucleotide geometry would not be accessible in

a duplex due to the additional constraints imposed by the neighbouring bases. One therefore has to ask whether or not there should be some evidence of this geometry in differences in fluorescence lifetime measurements. If the stability of the T-shaped geometry is similar to that of the duplex-like geometry then this might mean that a significant population of dinucleotides occupy this conformation in an ensemble. Furthermore, if this T-shaped population has different photophysical properties to the duplex form then this should be evident in fluorescence measurements.

Unfortunately this study cannot provide much insight into this issue and a much more rigorous investigation is required into the properties of the T-shaped geometry. For instance, despite the vastly different geometry, the photophysical properties of the T-shaped geometry might not be very different to those of the canonical form, although this is unlikely given the importance of charge transfer in this system. Additionally, while the structure might be relatively stable, it is possible that the energy barrier to achieving the T-shaped geometry is prohibitively high meaning that its occupancy would be relatively low. The ability of a dinucleotide to assume a T-shaped conformation may therefore be irrelevant in terms of fluorescence measurements.

5.4 Conclusions

The main objective of this investigation was to determine whether or not 2AP-containing dinucleotides were capable of adopting stacked structures similar those of B-form DNA. It was therefore necessary to begin the study by obtaining a good understanding of the parameter-space that was appropriate for DNA structure. While the archetypal double-helical structure is commonly described in very homogenous terms it was immediately evident that there is significant variation in the conformational properties of a real duplex. This is hardly new knowledge but the parameter values obtained by studying typical X-ray crystal structures of DNA provided a guide by which the computational results could be judged. It was also important to determine how appropriate the x3DNA analysis programs were for dinucleotide (or dimer) structures. The programs (as their names suggest) are designed to analyse DNA (particularly double-stranded DNA) and so it was necessary to ensure that they would give sensible values for such small constructs. An initial evaluation of the stacking parameters obtained for dinucleotides suggested that x3DNA programs would be suitable for this study. A major benefit of using the x3DNA programs was their use of standard definitions of DNA structure. Many previous studies have used schemes that are intuitive and easy to understand in isolation but, as they generally all differ to each other, it is difficult to relate the results of one study to another. Some further notes (including some potential disadvantages and shortcomings) about the use of x3DNA programs are given in Appendix 5, §5.5.

5.4.1 Dinucleotide Base-Stacking is Similar to DNA Duplex Conformation

It is clear from the results of this investigation that dinucleotides are able to form stable, stacked structures in both solvated and gas-phase environments. This demonstrates that the stacking interaction between only two bases is sufficiently favourable to hold a dinucleotide in a duplex-like conformation. There are some differences between the stacking geometry of the two constructs though. Perhaps the most obvious of these differences is found when comparing the twist angle between the bases. Dinucleotides, in general, were found to have a significantly larger twist angle than the value of 36° typically quoted for DNA. The same effect was seen in both 2AP dinucleotides and equivalent adenine dinucleotides. Therefore this perturbation in twist angle cannot simply be attributed to the use of an unnatural base. Comparison

with the results obtained for the dimer structures (where there was no constraint from the backbone) seems to suggest that the large twist values are due to the preferred stacking orientation of the bases. Each sequential base-step needs to be optimised in a duplex and so there is perhaps a balance between optimisation of local and global structure. The lack of base-pairing is also likely to be a significant cause of the discrepancy between twist values of dinucleotide and DNA structures. Investigation of extended structures might be able to reveal the exact reasons for the observed differences. It would also be interesting to determine the stacking conformation of the bases found at the end of duplex sequences. The results found here suggest that the difference between terminal and central bases is perhaps not as significant as might be expected. Therefore studies which utilise the ends of a duplex to investigate stacking energy are perhaps using a model which should give fairly accurate results.

5.4.2 Structure and 2AP Fluorescence Lifetime

The determination of 2AP-containing dinucleotide conformations which are comparable to those found in a duplex potentially has important consequences for the interpretation of fluorescence lifetime measurements. Specifically, the inference that the similarity of the fluorescence decay character in these constructs is due (at least partly) to structural similarities is corroborated by the results of this investigation.

An efficiently base-stacked dinucleotide conformation would seem to be a suitable mimic of the highly organised structure of the double helix. It can be inferred that this dinucleotide conformation (and similar conformations) is associated with the short-lifetime component observed in fluorescence lifetime measurements.

The conformation of base-stacked 2AP-N dinucleotides was found to be fairly independent of the natural base. There was a slight split between 2AP-R and 2AP-Y dinucleotides but the general structures were similar. It would therefore seem unlikely that conformation plays a major role in the differences observed in the short lifetime component of the fluorescence decay of these systems. The sequence-specific nature of the precise character of the short lifetime decay must therefore be mainly influenced by the intrinsic properties of the neighbouring base, such as its redox potential. It is possible that the strength of the stacking interaction might also play a role in determining the precise photophysical properties of the system.

The presence of a stable, open dinucleotide would (just like the base-flipped state in a duplex) be a plausible source for the longest lifetime component of the fluorescence decay. In this open dinucleotide conformation 2AP appears to be liberated

from intramolecular interactions and thus would be expected to behave as if it were essentially free in solution.

5.4.3 2AP as an Adenine Analogue

2AP-N dinucleotides and A-N dinucleotides share very similar overall geometries but there are some differences which could be of significance. The ability (and apparent preference) of 2AP to form HB1 might help to explain the results of Dallmann *et al.* which show that the 2AP·T base pair has a greater exchange rate than that of A·T.[74] Formation of HB1 would hinder base-pairing and thus lead to a greater base-flipping rate. The perturbation caused to the 2AP twist angle might also have a knock-on effect on the neighbouring bases as they would presumably rearrange to compensate for the loss of stacking efficiency. This could explain why the exchange rates of neighbouring base pairs were also affected by the change of a single adenine base to 2AP in the Dallmann *et al.* study. It would be interesting to compare the relative interaction energy of HB1 against 2AP·T base-pairing as this could provide insight into whether or not this hypothesis has any standing.

Adenine structures were found to be more stable than analogous 2AP structures; however, this could be attributed to the lower energy of the free adenine nucleobase compared to the free 2AP nucleobases.

Interestingly, despite the similarity of the dinucleotides, the stacking of 2AP|N dimers was found to be quite different to A|N dimers (except in the case of 2AP|A and A|A). It appears that without the constraint of the backbone the position of the NH_2 group plays a much more important role in determining the optimal geometry of the structure. It would seem plausible that this is an electrostatic effect.

5.4.4 The Sugar-Phosphate Backbone Influences Conformation

The importance of the sugar-phosphate backbone in determining the structure of nucleic acids is becoming ever more evident.[17] Although increasing the computational cost, inclusion of the sugar-phosphate backbone in the dinucleotide calculations meant that the structures obtained were constrained in a manner which was more akin to the natural system.

The results show that the backbone can influence the optimal interbase geometry. Comparison of analogous dimer and dinucleotide structures suggests that, in some cases, the twist angle is restricted by the presence of the backbone. Additionally, despite distorting base-stacking, backbone interactions (such as HB1 and HB2) were

found to be able to stabilise the dinucleotide structure. Structures containing HB2 were also observed in molecular dynamics simulations and were used to explain gas-phase, ion-mobility mass spectrometry results.[75]

It is worth noting that formation of HB1 could potentially influence the fluorescence properties of 2AP when incorporated into nucleic acid structure. Solvent H-bonding ability has been implicated as an important factor in determining the relaxation channel exploited by excited-state 2AP;[76] it would seem plausible that HB1 could have similar influence. A study by Lobsiger *et al.* suggests that the very nature of HB1 (where the H-bond actually occurs) might dramatically affect the observed 2AP photophysics.[77] It is also possible that HB1 could alter the electronic structure of 2AP by promoting pyramidalisation of the amino group.[18] It could even turn out that simply the increase in twist angle caused by the formation of HB1 is enough to sufficiently distort base-stacking such that quenching efficiency is reduced; for instance, electron transfer within DNA is known to be sensitive to perturbation of the base-stacked structure.[78] Although highly speculative, HB1 might potentially provide a gating mechanism if relaxation is reduced compared to more conventional base-paired conformations. Importantly, this could help to explain the presence of intermediate 2AP fluorescence lifetimes. It would therefore seem worthwhile investigating the photophysical properties of conformations containing HB1 in the future.

The specific structure of the backbone has not been considered in much detail during this investigation but there is perhaps useful information to be gained by doing so. A recent meta-study determined a number of distinct conformational families in DNA backbone structure.[79] It would be interesting to see if the backbone of a dinucleotide also conformed to such a description. It is possible that, due to the increased conformational freedom of the smaller system, the backbone of a dinucleotide could undertake a number of unique conformations.

There was concern that the charged phosphate group in the backbone might have had an influence on the obtained dinucleotide conformation; however, comparison of anionic and capped structures showed little difference in geometric parameters. Although there was a more significant change when the 5'-terminal phosphate group was included in calculations, particularly when PCM was excluded, it was concluded that using anionic dinucleotides would be suitable for this study.

While the charged phosphate group might not have influenced geometric structure it is possible that the presence of localised charge might have perturbed the photophysical properties of the system. A much more rigorous study would be required

to ascertain the validity of this proposition. A more advanced study might also be able to include the effect of counter-ions. A study by Churchill and Wetmore showed that counter-ions can have influence on the observed dinucleotide structure.[14]

5.4.5 PCM Solvent Effects

In general there was little difference observed between the geometry of base-stacked dinucleotides optimised in water or the gas-phase. Perhaps the most important observation was the change in the relative stability of the 2AP-T conformations when PCM solvent was excluded. This highlighted that different intra-molecular interactions became more or less important depending on the environment. In this particular case formation of HB2 in the gas-phase became more favourable than maintaining the base-stacking interaction which had been optimised in water.

Water and gas-phase dimer structures showed greater geometric differences than equivalent dinucleotides. It would appear that, without the backbone to stabilise the base-stacking orientation, the dimer structure rearranges to optimise electrostatic interactions.

Exclusion of PCM solvent also had more of an effect on noncanonical structures. For example, the T-shaped and open (flip2AP-T) geometries obtained for 2AP-T underwent significant structural rearrangement when the PCM solvent was removed. Solvent stabilisation clearly plays an important role in systems where the bases are more solvent exposed and isolated from intramolecular interactions.

It is worth mentioning that calculations were also performed using a variety of different PCM solvents other than water. The results of these calculations are discussed in more detail in Appendix 5, §5.2 but the main point of note is the fact that a systematic change was observed when altering the PCM solvent. While the dielectric coefficient is not the only parameter that changes between solvents it appeared to have a significant bearing on the energy of the optimised structure and a very simple linear trend was observed. This suggests that only the interaction strength and not the specific interactions were changing with solvent. This matches well with the fact that the geometry of the dinucleotides remained essentially constant (implying that similar interactions would be present).

5.4.6 Caveats and Possible Future Experiments

There are, of course, some caveats to the interpretation of the results of this investigation which are worth bearing in mind.

Using PCM for solvation has limitations (for example, solute-solvent H-bonding is not taken into account) and future studies might benefit from the use of explicit solvent models. That being said, it is encouraging to see that PCM appears to at least give sensible looking results as using an explicit model would come at increased computational cost.

There is almost always the potential to use a more advanced functional or larger basis set to improve results but, again, the gains have to be weighed up against the additional time it takes to obtain the results.

Unfortunately gradient optimisation methods (such as the one utilised here) are susceptible to getting stuck in shallow minima and so the optimised structure obtained is sensitive to the starting geometry used. As many energy minima are possible for the systems studied in this investigation it is highly unlikely that global minima structures were found. For example, the formation of HB2 in 2AP-T dinucleotides pulled the structure away from a more stable form where base-stacking was better optimised (see §5.3.5). Comparison of optimised energies is therefore of limited quantitative value unless a more extensive exploration of the potential energy surface is performed.

This investigation has perhaps provided more questions than answers. The vast parameter-space in which nucleic acid constructs exist means that there are seemingly endless avenues to explore. For example, it would be beneficial to study base-stacking for all possible permutations of bases but, just considering canonical bases, the required number of calculations quickly becomes prohibitively time-consuming. With only 4 natural bases there are 16 (4^2) possible base steps (10 of which are unique if it is assumed base-pairing results in identical parameters for complementary doublets; for example, 5'-A-C-3' would be equivalent to 5'-G-T-3'). Although computationally expensive, recent studies are beginning to show that such an exhaustive approach is becoming feasible.[50,80] This being said, there is likely considerable value in focussing efforts on a limited number of experiments that can elucidate upon particular problems. Some examples are given below.

While it has been shown that dinucleotides can assume base-stacked and open conformations this study has not provided any insight into the way in which the dinucleotide actually transitions between these two extremes. As this process is

analogous to base-flipping it would be very useful to be able to gain knowledge about this pathway. There have been a number of studies which have utilised molecular dynamics to this effect but, unfortunately, these methods generally have reduced accuracy and require restraints to be imposed upon the system.[61] Therefore, although they provide an overall picture of the process, a more rigorous approach is required to gain a detailed view of what is happening.

Investigation into the mechanism of base-flipping will be continued as part of the collaboration with Tanja van Mourik and Leo Holroyd (University of St Andrews, St Andrews, United Kingdom). An initial approach will be to try and obtain stable dinucleotide structures that have conformations somewhere between the stacked and open geometries. An attempt will then be made to try and link these structures to obtain a sequence that connects the two extremes. The hope is that this will help to elucidate the base-flipping process by providing potential intermediate states. In addition, the presence of these conformations might give credence to the idea that the intermediate lifetimes observed during time-resolved fluorescence measurements are due to structures in which the bases are neither fully stacked nor fully isolated from each other. It will also be of interest to see how the energy of the structure changes as the dinucleotide opens out from its stacked (and presumably most stable) state. For this comparison of energies to be meaningful it will be important to ensure that the global minimum has been found for the base-stacked structures. This could be a significant challenge in itself given the large conformational space which must be searched. One approach might be to search for plausible conformations using molecular mechanics (a fast but low accuracy method) and then refine the resulting structures with a high accuracy quantum method.[81–83]

It is interesting that the Hoogsteen and T-shaped 2AP-T conformations were apparently of similar stability to the canonical base-stacked structure. There is therefore perhaps value in investigating these unusual structures in more detail to determine if they would be significantly populated within an ensemble. If such conformations have distinct photophysical properties then they might contribute to the complicated nature of the fluorescence of 2AP incorporated into nucleic acids. They could also possibly explain subtle differences between the fluorescence decay of dinucleotides and larger constructs as certain conformations might only be accessible in the relatively flexible dinucleotide.

A potentially straight-forward experiment to perform would be to calculate the stability of dinucleotide conformations when starting from typical A-DNA form (or,

indeed, any other known DNA conformation). It would be valuable to know if the dinucleotide structure reverted back to B-DNA form or if it maintained a unique conformation. Although B-DNA is the most common DNA structure it is possible that other conformations are more stable for a relatively simple dinucleotide.

This investigation has, somewhat arbitrarily, used dinucleotides of the form 5'-2AP-N-3' but it would perhaps be worthwhile performing calculations on the reversed sequence (5'-N-2AP-3'). This would provide insight into whether or not the order of the bases influences the dinucleotide conformation. Quenching *via* charge transfer within a duplex has been found to be dependent on the position (5' or 3') of the 2AP relative to a proximal guanine base.[84] It would be interesting to see if conformation played any role in this difference.

A related study would be to look at the stacking of 2AP with inosine (a nucleotide with hypoxanthine as the base moiety). Despite hypoxanthine being structurally similar to guanine (lacking only the C2 NH_2 group) the redox potential of inosine is thought to be prohibitive to charge transfer quenching of 2AP. Again, it would be useful to see if conformation had any influence on the differences observed.

2D-fluorescence spectroscopy was recently used to investigate a 2AP-2AP dinucleotide.[85] The results suggested that the 2AP bases stack with a twist angle of only 5°. There is clearly scope for performing calculations to investigate if such a conformation is stable for this dinucleotide.

An obvious progression of this investigation is the extension of the nucleic acid structures. This could be achieved by increasing the number of nucleotide units in a single-strand or by introducing a complementary base-pairing strand (or both). There are a number of questions which could be addressed by studying these larger constructs: Is the inclusion of only a second neighbouring base enough to cause a more duplex-like stacking conformation? Does base-pairing limit the twist angle to a more restricted range? What backbone interactions are possible within a longer nucleic acid sequence? As with this investigation, there are also likely to be many questions that are currently unknowable and will only arise once initial calculations on these systems are performed. Unfortunately exponential growth of the computational cost with system size is a significant barrier to these experiments; however, calculations on larger nucleic acid structures are starting to be performed.[16,50]

5.4.7 Outlook

Computational studies are likely to be at the forefront of any further advances that are made to the understanding of nucleic acid structure. A major advantage of these methods is that it is possible to alter many parameters very easily. For example, the ability to perform calculation with and without PCM solvent provided a convenient means to compare solvated and gas-phase structures. The fact it was simple to modify the dinucleotide structures also meant that a more thorough investigation could be performed. In particular, comparison between 2AP and adenine dinucleotides was viable because the NH_2 group position could be altered at will. The ever increasing computational power available to researchers will continue to expand the versatile nature of these methods; systems that once seemed complex will soon become trivial. That being said, the reliability of computational methods can really only be vetted by the results of practical experiments. There is a distinct lack of much needed literature to confirm or reject the recent explosion of theoretical predictions made by computational studies. A particularly relevant omission in regard to this study is the fact that there are few estimates of the energy associated with base-stacking. Estimates that do exist have generally considered large (duplex) systems where specific base-base interactions are difficult to isolate from other contributing factors such as solvation.[86–88] Some confidence can be gained from the fact that other parameters of these systems can be predicted with good accuracy. For example, the energies calculated for base-pairing interactions matches well with experiment. This is encouraging but it is clear that there is still significant room for improvement.

The importance of DNA is undeniable but its complexity means that it is also, simply, a fascinating system to study. As more knowledge is gained more questions seem to arise. Even though it has been 60 years since Watson and Crick solved the structure of DNA it would appear that it is going to be a very long time before all of its secrets are discovered. The picture of DNA as a static scaffold is not sufficient and the importance of its dynamic properties are becoming ever more evident. In the past the study of smaller dinucleotide systems led to advances in the understanding of DNA structure by helping to confirm base-pairing. It would seem likely that a similar approach will assist in determining the dynamic properties of DNA. While this study only considered isolated and static dinucleotide structures it provides a framework by which transitional studies can be performed. For example, base-flipping can be investigated by using the stacked and open dinucleotide conformations as possible

start- and end-points. As this mechanism is known to be important in a number of biological functions it is highly desirable to know how it transpires.[89] This investigation also provides some insight into why the photophysical behaviour of dinucleotides is similar to much larger constructs. Specifically, it has been found that 2AP-N dinucleotides can assume stable conformations comparable to those found in a duplex. Similar studies should allow even greater insight into the complex world of nucleic acids; a world that it is essential we understand.

5.5 References

1. W. C. Johnson, Jr. and I. Tinoco, Jr., *Biopolymers*, 1969, **7**, 727–49.

2. C. R. Cantor, M. M. Warshaw, and H. Shapiro, *Biopolymers*, 1970, **9**, 1059–77.

3. N. P. Johnson and E. Switkes, *Biopolymers*, 1978, **17**, 857–72.

4. D. M. Cheng and R. H. Sarma, *J. Am. Chem. Soc.*, 1977, **99**, 7333–48.

5. F. S. Ezra, C.-H. Lee, N. S. Kondo, S. S. Danyluk, and R. H. Sarma, *Biochemistry*, 1977, **16**, 1977–87.

6. P. M. Keane, M. Wojdyla, G. W. Doorley, J. M. Kelly, I. P. Clark, A. W. Parker, G. M. Greetham, M. Towrie, L. M. Magno, and S. J. Quinn, *Phys. Chem. Chem. Phys.*, 2012, **14**, 6307–11.

7. G. W. Doorley, M. Wojdyla, G. W. Watson, M. Towrie, A. W. Parker, J. M. Kelly, and S. J. Quinn, *J. Phys. Chem. Lett.*, 2013, **4**, 2739–44.

8. J. Šponer, P. Jurečka, and P. Hobza, in *Computational Studies of RNA and DNA*, eds. J. Šponer and F. Lankaš, Springer, 2006, pp. 343–88.

9. A. Abo-Riziq, L. Grace, E. Nir, M. Kabelac, P. Hobza, and M. S. de Vries, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 20–3.

10. E. Nir, K. Kleinermanns, and M. S. de Vries, *Nature*, 2000, **408**, 949–51.

11. M. Kabeláč, F. Ryjáček, and P. Hobza, *Phys. Chem. Chem. Phys.*, 2000, **2**, 4906–9.

12. C. Fonseca Guerra, F. M. Bickelhaupt, J. G. Snijders, and E. J. Baerends, *J. Am. Chem. Soc.*, 2000, **122**, 4117–28.

13. R. S. Hunter and T. van Mourik, *J. Comput. Chem.*, 2012, **33**, 2161–72.

14. C. D. M. Churchill and S. D. Wetmore, *Phys. Chem. Chem. Phys.*, 2011, **13**, 16373–83.

15. V. I. Poltev, V. M. Anisimov, V. I. Danilov, D. Garcia, A. Deriabina, E. González, R. Salazar, F. Rivas, and N. Polteva, *Comput. Theor. Chem.*, 2011, **975**, 69–75.

16. T. A. Zubatiuk, O. V. Shishkin, L. Gorb, D. M. Hovorun, and J. Leszczynski, *Phys. Chem. Chem. Phys.*, 2013, **15**, 18155–66.

17. V. Poltev, V. M. Anisimov, V. I. Danilov, D. Garcia, C. Sanchez, A. Deriabina, E. Gonzalez, F. Rivas, and N. Polteva, *Biopolymers*, 2013, **101**, 640–50.

18. J. M. Jean and K. B. Hall, *J. Phys. Chem. A*, 2000, **104**, 1930–7.

19. J. M. Jean and K. B. Hall, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 37–41.

20. J. M. Jean and K. B. Hall, *Biochemistry*, 2002, **41**, 13152–61.

21. S. J. O. Hardman and K. C. Thompson, *Biochemistry*, 2006, **45**, 9145–55.

22. J. Liang and S. Matsika, *J. Am. Chem. Soc.*, 2011, **133**, 6799–808.

23. J. Liang, Q. L. Nguyen, and S. Matsika, *Photochem. Photobiol. Sci.*, 2013, **12**, 1387–400.

24. T. van Mourik, personal communication, School of Chemistry, University of St Andrews, St Andrews, United Kingdom, 2013.

25. T. Lenz, E. Y. M. Bonnist, G. Pljevaljčić, R. K. Neely, D. T. F. Dryden, A. J. Scheidig, A. C. Jones, and E. Weinhold, *J. Am. Chem. Soc.*, 2007, **129**, 6240–8.

26. K. Hoogsteen, *Acta Crystallogr.*, 1959, **12**, 822–3.

27. K. Hoogsteen, *Acta Crystallogr.*, 1963, **16**, 907–16.

28. P. Várnai and K. Zakrzewska, *Nucleic Acids Res.*, 2004, **32**, 4269–80.

29. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, *Gaussian 09, Revision A.02, Gaussian, Inc., Wallingford CT*, 2009.

30. Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215–41.

31. A. Herráez, *Biochem. Mol. Biol. Educ.*, 2006, **34**, 255–61.

32. Jmol: an open-source Java viewer for chemical structures in 3D, *Version 13.0.8*, See; http://www.jmol.org/, 2012.

33. The PyMOL Molecular Graphics System, *Version 1.6.0.0*, Schrödinger, LLC, See; http://www.pymol.org/, 2013.

34. G. Schaftenaar and J. H. Noordik, *J. Comput. Aided. Mol. Des.*, 2000, **14**, 123–34.

35. MOLDEN: a pre- and post-processing program of molecular and electronic structure, *Version 5.0*, See; http://www.cmbi.ru.nl/molden/, 2010.

36. X.-J. Lu and W. K. Olson, *Nucleic Acids Res.*, 2003, **31**, 5108–21.

37. X.-J. Lu and W. K. Olson, *Nat. Protoc.*, 2008, **3**, 1213–27.

38. G. Zheng, X.-J. Lu, and W. K. Olson, *Nucleic Acids Res.*, 2009, **37**, W240–6.

39. T. van Mourik, personal communication, School of Chemistry, University of St Andrews, St Andrews, United Kingdom, 2013.

40. MATLAB, *Version 8.2.0 (R2013b)*, The MathWorks, Inc., Natick, Massachusetts, United States of America, 2013.

41. C. Altona and M. Sundaralingam, *J. Am. Chem. Soc.*, 1972, **94**, 8205–12.

42. M. A. El Hassan and C. R. Calladine, *J. Mol. Biol.*, 1995, **251**, 648–64.

43. M. A. El Hassan and C. R. Calladine, *Philos. Trans. R. Soc. London. Ser. A*, 1997, 43–100.

44. IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN), *Eur. J. Biochem.*, 1983, **131**, 9–15.

45. N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, *J. Cheminform.*, 2011, **3**, 33.

46. SketchUp, *Version 8.0*, Trimble Navigation Ltd., Sunnyvale, California, United States of America, See; http://www.sketchup.com/, 2012.

47. W. K. Olson, M. Bansal, S. K. Burley, R. E. Dickerson, M. Gerstein, S. C. Harvey, U. Heinemann, X.-J. Lu, S. Neidle, Z. Shakked, H. Sklenar, M. Suzuki, C.-S. Tung, E. Westof, C. Wolberger, and H. M. Berman, *J. Mol. Biol.*, 2001, **313**, 229–37.

48. P. Hobza, H. L. Selzle, and E. W. Schlag, *J. Phys. Chem.*, 1996, **100**, 18790–4.

49. W. Kabsch, C. Sander, and E. N. Trifonov, *Nucleic Acids Res.*, 1982, **10**, 1097–104.

50. G. Barone, C. Fonseca Guerra, and F. M. Bickelhaupt, *ChemistryOpen*, 2013, **2**, 186–93.

51. K. R. Julien, M. Sumita, P.-H. Chen, I. A. Laird-Offringa, and C. G. Hoogstraten, *RNA*, 2008, **14**, 1632–43.

52. L. Holroyd, personal communication, School of Chemistry, University of St Andrews, St Andrews, United Kingdom, 2013.

53. N. Tibanyenda, S. H. de Bruin, C. A. G. Haasnoot, G. A. van der Marel, J. H. van Boom, and C. W. Hilbers, *Eur. J. Biochem.*, 1984, **27**, 19–27.

54. G. Altan-Bonnet, A. Libchaber, and O. Krichevsky, *Phys. Rev. Lett.*, 2003, **90**, 138101.

55. T. M. Nordlund, S. Andersson, L. Nilsson, R. Rigler, A. Gräslund, and L. W. McLaughlin, *Biochemistry*, 1989, **28**, 9095–103.

56. D. Xu, K. O. Evans, and T. M. Nordlund, *Biochemistry*, 1994, **33**, 9592–9.

57. M. Guéron, M. Kochoyan, and J.-L. Leroy, *Nature*, 1987, **328**, 89–92.

58. M. Guéron and J.-L. Leroy, *Methods Enzymol.*, 1995, **261**, 383–413.

59. P. K. Bhattacharya, J. Cha, and J. K. Barton, *Nucleic Acids Res.*, 2002, **30**, 4740–50.

60. U. Dornberger, M. Leijon, and H. Fritzsche, *J. Biol. Chem.*, 1999, **274**, 6957–62.

61. U. D. Priyakumar and A. D. MacKerell, Jr., *Chem. Rev.*, 2006, **106**, 489–505.

62. P. H. von Hippel, N. P. Johnson, and A. H. Marcus, *Biopolymers*, 2013, **99**, 923–54.

63. D. Coman and I. M. Russu, *Biophys. J.*, 2005, **89**, 3285–92.

64. E. Giudice, P. Várnai, and R. Lavery, *Chemphyschem*, 2001, **2**, 673–7.

65. P. Várnai and R. Lavery, *J. Am. Chem. Soc.*, 2002, **124**, 7272–3.

66. E. Giudice, P. Várnai, and R. Lavery, *Nucleic Acids Res.*, 2003, **31**, 1434–43.

67. U. D. Priyakumar and A. D. MacKerell, Jr., *J. Am. Chem. Soc.*, 2006, **128**, 678–79.

68. S. Klimasauskas, S. Kumar, R. J. Roberts, and X. Cheng, *Cell*, 1994, **76**, 357–69.

69. M. O. Sinnokrot and C. D. Sherrill, *J. Phys. Chem. A*, 2004, **108**, 10200–7.

70. P. Linse, *J. Am. Chem. Soc.*, 1992, **114**, 4366–73.

71. R. A. DiStasio, Jr., G. von Helden, R. P. Steele, and M. Head-Gordon, *Chem. Phys. Lett.*, 2007, **437**, 277–83.

72. C. M. Baker and G. H. Grant, *J. Chem. Theory Comput.*, 2006, **2**, 947–55.

73. J. G. Hill, J. A. Platts, and H.-J. Werner, *Phys. Chem. Chem. Phys.*, 2006, **8**, 4072–8.

74. A. Dallmann, L. Dehmel, T. Peters, C. Mügge, C. Griesinger, J. Tuma, and N. P. Ernsting, *Angew. Chem. Int. Ed. Engl.*, 2010, **49**, 5989–92.

75. J. Gidden and M. T. Bowers, *Eur. Phys. J. D*, 2002, **20**, 409–19.

76. C. Reichardt, C. Wen, R. A. Vogt, and C. E. Crespo-Hernández, *Photochem. Photobiol. Sci.*, 2013, **12**, 1341–50.

77. S. Lobsiger, S. Blaser, R. K. Sinha, H.-M. Frey, and S. Leutwyler, *Nat. Chem.*, 2014, **6**, 989–93.

78. S. O. Kelley and J. K. Barton, *Science*, 1999, **283**, 375–81.

79. D. Svozil, J. Kalina, M. Omelka, and B. Schneider, *Nucleic Acids Res.*, 2008, **36**, 3690–706.

80. V. I. Poltev, V. M. Anisimov, V. I. Danilov, T. van Mourik, A. Deriabina, M. Padua, D. Garcia, E. Gonza, F. Rivas, and N. Polteva, *Int. J. Quantum Chem.*, 2010, **110**, 2548–59.

81. D. C. Young, *Computational Chemistry - A Practical Guide for Applying Techniques to Real World Problems*, John Wiley & Sons, Inc., 2001.

82. F. Jensen, *Introduction to Computational Chemistry*, John Wiley & Sons, Ltd, 2nd edn., 2007.

83. R. Galindo-Murillo, C. Bergonzo, and T. E. Cheatham, III, *Curr. Protoc. Nucleic Acid Chem.*, 2013, **54**, 7.5.1–7.5.13.

84. M. A. O'Neill and J. K. Barton, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 16543–50.

85. J. R. Widom, N. P. Johnson, P. H. von Hippel, and A. H. Marcus, *New J. Phys.*, 2013, **15**, 025028.

86. J. SantaLucia, Jr., *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 1460–5.

87. E. Protozanova, P. Yakovchuk, and M. D. Frank-Kamenetskii, *J. Mol. Biol.*, 2004, **342**, 775–85.

88. P. Yakovchuk, E. Protozanova, and M. D. Frank-Kamenetskii, *Nucleic Acids Res.*, 2006, **34**, 564–74.

89. R. J. Roberts and X. Cheng, *Annu. Rev. Biochem.*, 1998, **67**, 181–98.

Chapter 6: Reversible Fluorescence Photoswitching in DNA

We see a lot of beautiful things through the microscope and the telescope.

– Thomas Ebbesen, *ICP 2013* (KU Leuven)

This chapter is based on published work from D. A. Smith, P. Holliger, and C. Flors, *J. Phys. Chem. B*, 2012, **116**, 10290–3.

6.1 Introduction

Fluorescence photoswitching plays a fundamental role in a number of optical microscopy techniques that are capable of surpassing the spatial resolution conventionally limited by the diffraction of light, the so-called super-resolution techniques.[1,2] The desire to image systems on an ever smaller spatial scale has therefore led to considerable advances in fluorescence photoswitching in recent years.

The rapid growth of knowledge and technologies within this field has also enabled the development of novel techniques, such as optical lock-in detection (OLID) microscopy,[3–5] which are capable of improving dynamic contrast within an image.[6] Rather than improving spatial resolution, OLID microscopy enhances the image contrast (signal-to-noise ratio) by utilising optical control to modulate the fluorescence emission of the target. Cross-correlation with a reference waveform subsequently allows discrimination of the signal of interest from the unmodulated background.

The remarkable progress that has been made in imaging capability has offered great promise in improving our understanding of the sub-microscopic realm. For instance, the structure of the biological cell has long been studied through the microscope but conventional methods have been limited in the level of detail they could capture. Super-resolution microscopy could therefore allow access to a previously unseen cellular world.[7] This potential will, however, only come to fruition with the development of strategies capable of labelling cell components with the necessary properties required for such imaging methods.

Since the best attainable resolution is primarily determined by label density, it is essential that the highest labelling density is achieved within a system. Through vigorous effort, there are now various methods available to utilise proteins as fluorescent labels for microscopy. For example, the discovery and subsequent engineering of fluorescent proteins has produced whole libraries of candidates for *in vivo* cellular imaging. In contrast, there are still few approaches that are able to exploit DNA in a similar manner. Super-resolution microscopy of DNA has, so far, only been possible using stochastic blinking or transient binding of noncovalently bound dyes,[8–12] click-chemistry,[13,14] and by the use of short, complementary docking and imager DNA strands.[15]

This chapter reports an attempt to apply a new approach to attain reversible fluorescence photoswitching in DNA. A recent study developed an elegant method to incorporate a high label density into a DNA substrate by utilising a modified DNA

polymerase.[16] Directed evolution produced a polymerase able to tolerate bulky Cy3- and Cy5-labelled cytidine nucleotide analogues and enabled efficient incorporation of a dense population of fluorophore labels within a duplex *via* polymerase chain reaction (PCR). The resulting DNA substrate, coined CyDNA, extended to around 1.3 kbp and each strand could be decorated with hundreds of cyanine fluorophores; however, despite being strongly coloured and exhibiting considerable fluorescence emission, early observations suggested that the high labelling density caused significant fluorescence quenching, particularly for the Cy5 variant.

One of the priorities of this investigation was to characterise the mechanisms of fluorescence quenching within CyDNA. Improved understanding of these pathways would allow optimisation of labelling strategies for this system in future studies.

Reversible fluorescence photoswitching was introduced into CyDNA by utilising the desirable properties of the Cy3-Cy5 pair.[17,18] Previous studies showed that, in the presence of a thiol and an oxygen scavenging system, proximal Cy3 and Cy5 dyes form an efficient photoswitch that can be activated at 532 nm and deactivated at 638 nm.[19]

The photoswitching capability of CyDNA has been investigated at ensemble and single-molecule scales. As the quality of an OLID image is highly dependent on the controllable nature of the photoswitching behaviour in the system, OLID microscopy served as an excellent tool to study the reliability of the fluorescence photoswitching in CyDNA. A proof-of-principle super-resolution imaging experiment was also performed to confirm the potential of CyDNA as a probe of nanoscale structure.

As further introduction, a more detailed description of some of the more important principles involved in this study will be presented below. Unless otherwise stated, the term *microscopy* is taken to mean optical microscopy throughout this chapter.

6.1.1 Advantages of Microscopy

Microscopy has become an invaluable tool in the scientific world; in particular, the understanding of biological systems has been greatly advanced with its development. The fact many processes in biology occur within the micron-millimetre scale has clearly been a significant factor in the success of microscopy within this field. There are, however, other important advantages of using this technique.

In principle, microscopy is completely non-invasive. This is particularly useful for biological systems, which are often significantly altered – even destroyed – by a foreign interaction. The non-destructive nature of microscopy means that it is possible to

perform *in vivo* studies and so obtain results most realistic to the natural, living system. Detection sensitivity is also an important consideration for a technique that is used to study systems that are inherently complex. Microscopy has been shown to be able to detect at the ultimate limit of sensitivity; single-molecule detection.[20–22]

Single-molecule studies can provide a much greater depth of knowledge about a system; the results from such studies avoid ensemble averaging. This means unsynchronised behaviour, such as competitive reaction pathways and transient states, can be studied.[23] This being said, single-molecule experiments are potentially time-consuming because a large number repeats is required to obtain a reasonable level of confidence in the results. They also sometimes require manipulation of the system, such as immobilisation of the target molecules; this contradicts the non-invasive premise of microscopy and so it can become unclear as to whether or not the data are accurate for a real system.[23]

A whole range of microscopy techniques has been developed over the years in an attempt to increase the level of detail which can be achieved, a selection of which is now briefly described.

6.1.2 Microscopy Techniques

Bright-field microscopy relies on absorption of light by the sample to produce contrast to an otherwise transmitted illumination source. Dark-field microscopy, which can almost be considered the conjugate partner of bright-field microscopy, increases image contrast by only collecting the light which is scattered by the sample.

Objects that would otherwise be transparent can be imaged by using phase shift[24–26] or interference microscopy techniques.[27,28]

Confocal microscopy improves image resolution and allows the 3-dimensional (3D) structure of a sample to be determined.[29] These advantages come as a result of collecting light from only a small, isolated spot (volume) at any given time and so the trade-off is that a raster scan has to be performed to construct the full image.

Wide-field techniques allow fast imaging across relatively large areas in the sample. This is advantageous because, in contrast to scanning techniques, the images produced are temporally synchronised without further processing. Wide-field techniques also avoid potential problems caused by having a scan-time, such as artefacts caused by the system changing on the scan timescale.

Fluorescence microscopy plays a fundamental role in the results reported in this chapter and so will be discussed in greater depth in the following section.

6.1.2.1 Fluorescence Microscopy

Fluorescence microscopy has become a vital tool for studying a diverse range of systems due to its sensitivity and versatility. Selective excitation of a fluorophore label means that it is possible to visualise a specific target, even within a complex environment. Additionally, by using a different wavelength of light for excitation and detection, scattered (excitation) light can be discarded and only signal from the target is detected. Epi-fluorescence techniques optimise collection from the excitation spot by detecting emission in the same direction as the excitation source; this helps to minimise signal losses. Total internal reflection fluorescence (TIRF) microscopy utilises the fact that evanescent waves are produced when light is reflected at a surface. Evanescent waves can be used to excite fluorophores to within around 100-200 nm of the surface. Thus, if the sample is known to be found on the surface, the background signal from bulk molecules can be reduced by using this method.[4]

Fluorescence microscopy is obviously reliant on the use of fluorescent probes. Successful imaging requires bright probes (high extinction coefficient and quantum yield) that are stable within the target environment and also resistant to photobleaching. The discovery (for example, green fluorescent protein, GFP[30]) and synthesis (for example, cyanine dyes[31,32]) of novel fluorescent materials with these properties has therefore been of great importance in the development of this technique.

Introduction of probes to a system is, however, potentially challenging. For example, delivery must not disrupt normal function if an accurate description of the system is to be obtained. Toxicity is a particular problem for biological studies. Exogenous probes must also target the specific process or structure under investigation otherwise they will increase the background signal.

The remainder of this chapter will mainly be concerned with fluorescence microscopy; however, some of the discussion which follows, such as the next section on resolution, is generally applicable to other microscopy methods.

6.1.3 Spatial Resolution and the Diffraction Limit

An ever increasing level of detail is sought by the scientific community.[33] It is therefore necessary for microscopy to overcome the challenges faced by the current level of technology. The most significant advance that has been made in recent years has been the development of techniques capable of improving spatial resolution beyond the diffraction limit.

The signal from a sample is affected by the various optical components used in a microscope and so the image of even an ideal point-source is distorted. This distortion is known as the point-spread-function (PSF) of the microscope. The observed signal from a sample is a convolution of the true signal and the PSF.

Diffraction of light is a significant contributor to the PSF of conventional optical microscopy techniques, limiting their optimal resolution to around 250-300 nm for the visible spectrum. Factors which contribute to this limit include the wavelength of light collected, λ, and the numerical aperture, NA, of the microscope used. The NA is defined by the refractive index of the imaging medium (usually air or oil), η, and the half-angle of the angular aperture, α (essentially the collecting power of the lens). Figure 6.1 shows a schematic of these parameters.

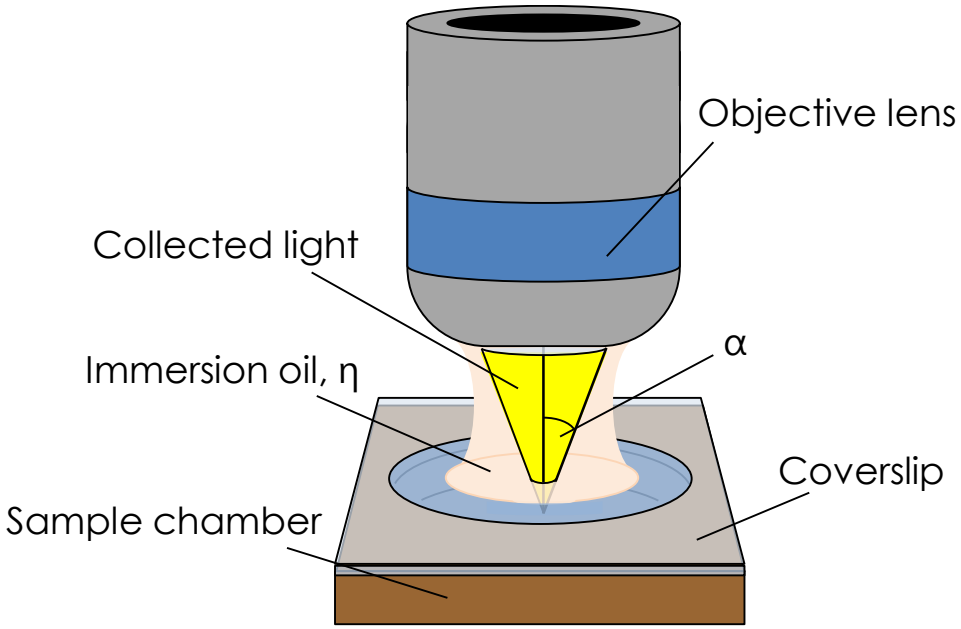

Figure 6.1 – Schematic of the parameters which define the diffraction-limited resolution of a microscope. If either the refractive index of the imaging medium, η, or the half-angle of the angular aperture, α, are increased then resolution is improved. For clarity the distance between the objective and the coverslip has been exaggerated and the rest of the microscope (for example, the camera detector) beyond the objective has been omitted.

The (spatial) resolution, R_c, of a system is often defined by the Rayleigh criterion (or Rayleigh limit);[34–36]

$$R_c = \frac{1.22\lambda}{2\eta \sin\alpha} = \frac{0.61\lambda}{NA}, \qquad 6.1$$

where the numerical factor 1.22 is a consequence of using a model based on the Airy (disc) pattern to define the discernible contrast limit. An Airy pattern is caused by the diffraction of light as it passes through a circular aperture (such as an objective lens) and is characterised by a central disc (Airy disc) surrounded by a series of concentric rings.

Figure 6.2 shows examples of the intensity distribution detected from two point-sources which are resolvable, at the Rayleigh limit, or unresolvable. At the Rayleigh limit the two peaks are a distance R_c apart.

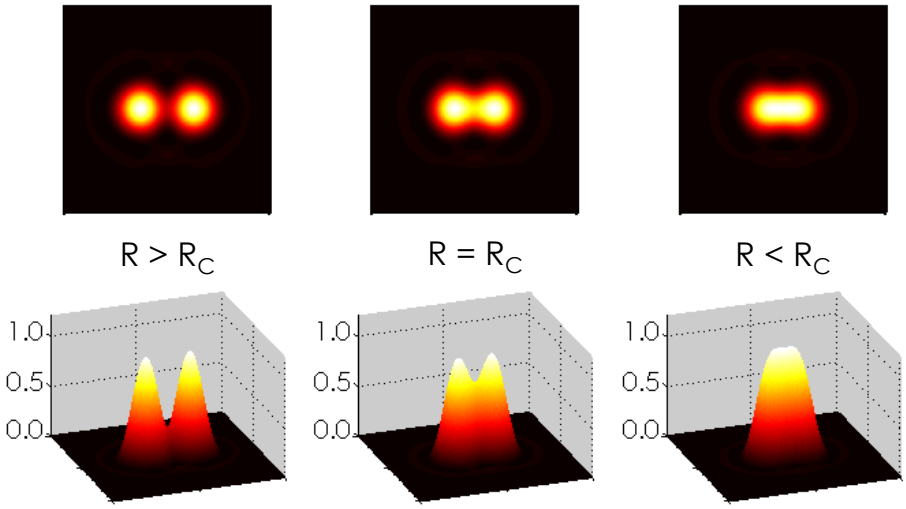

Figure 6.2 – Comparison of the measured signal (intensity distribution) from two point-sources, separated by distance R, which are resolved (left), at the Rayleigh criterion (middle), or unresolved (right). The top row shows contour plots while the bottom row shows perspective views. Intensity values range from 0 (black) to 1 (white). Spatial axes have arbitrary units but are scaled such that when R is equal to R_c the first minimum of each PSF is at the maximum of the other point-source.

It can be seen from Equation 6.1 that the resolution of a system can be improved by increasing the NA of the system or by decreasing the wavelength of light collected. Unfortunately the NA is limited to around 1.4-1.6 for current technology. Using light of a shorter wavelength is possible but the higher energy can result in undesirable effects such as damage to the sample. Additionally, Rayleigh scattering increases as the wavelength is reduced and so high background signal can become problematic. The wavelength of light might also be restricted by the sample being studied or the optical components which are used.

The need for higher resolution is apparent in many different areas, such as in the study of DNA topology[9] or the structure of mitochondrial cristae.[37] The drive for a resolution beyond the diffraction limit has resulted in the development of a multitude of elegant techniques capable of achieving this aim. Table 6.1 provides a very brief overview of some of the advantages and disadvantages of a selection of the more prominent methods currently used. A number of these techniques rely heavily on photoswitching, which will be discussed in the next section.

Technique	Advantage	Disadvantage
Atomic Force Microscopy (AFM)[38]	Very high resolution possible (Å). Visualise bonds.[39–41] Functionalise probe.	Only possible to image the surface of materials.
Scanning Tunnelling Microscopy (STM)[42,43]	Very high resolution possible (Å). Manipulation of system.[44] Spectroscopy (STS).	Limited to surface measurements. Lack of chemical sensitivity.
Scanning Electron Microscopy (SEM)[45,46]	Very high resolution possible (Å).	Require high vacuum. Sample surface must be conductive or coated.
Superlenses[47]	White-light illumination.[48] Fluorophores not required.	Lack selectivity of fluorescence. Contrast limited.
Near-Field Scanning Optical Microscopy (NSOM, SNOM)[49–51]	Selective detection though fluorescence. Topological information.	Relatively poor resolution enhancement for additional set-up time and complexity.
Structured Illumination (SIM/SSIM/NL-SIM)[52–54]	Fast acquisition. Potential for single-shot measurements.[55]	Limited to only twice diffraction limited resolution unless high laser power used.
Stimulated Emission Depletion (STED)[56,57]	High resolution in 3D. Relatively short acquisition times. Parallelisation possible.[58]	Can require high laser power and specialised optics.
Single-Molecule Localisation Microscopy (SMLM)[2] Photoactivated Localisation Microscopy (PALM/fPALM)[59–61] Stochastic Optical Reconstruction Microscopy (STORM/dSTORM)[17,62]	Fairly simple set-up. Standard fluorophores can be used.[63] 3D imaging.[64,65]	Long acquisition times. Requires accurate fluorophore localisation and image reconstruction.[66]

Table 6.1 – A selection of super-resolution microscopy techniques capable of surpassing the diffraction limit. PALM and STORM can both be described as SMLM techniques. STED and SMLM require the use of molecules which can photoswitch between distinct photophysical states. AFM, STM, and SEM are not optical techniques but are provided for comparison.

6.1.4 Optical Control: Photoswitching

There are many situations in which light can be used to induce some change in the properties of a system. Such a change can be categorised under the general term photoconversion. Of particular relevance to this study is the alteration of the photophysical properties of the target by light.

Photoactivation refers to a change from a non-detectable, dark state to a detectable state. The term *dark* does not necessarily mean that the molecule is non-emissive; it simply means that the target cannot be detected by the imaging method used or, perhaps, is deliberately rejected through the use of a filter. Photoactivation specifically refers to the transition of a target to its active (that is, measurable) state but it may lead to a whole chain of further events or be part of a reversible process, such as photoswitching, which is discussed below. There are a vast number of molecules that have been identified as suitable candidates for photoactivation investigations. Colourless spiropyrans and spirooxazines have been studied for decades because of their ability to form coloured merocyanines.[67–69] Organic fluorophores, such as Alexa or ATTO dyes, and fluorescent proteins, such as Kaede,[70] kindling fluorescent protein-1 (KFP-1),[71] and photoactivatable GFP[72] have more recently been developed to expand the versatility of photoactivatable probes.[73,74]

Photochromism is where irradiation causes a change in absorption and/or emission spectra. This can be irreversible, as in Kaede, or reversible, as in spiropyrans and spirooxazines. Photochromism has found application in photosensitive glasses.

Photoswitching is a general term for a two-state system (for example, emissive and dark) in which states can be reversibly accessed using light (in at least one direction), as in Dronpa.[75] Importantly, the photophysical properties of a photoswitching molecule can be optically controlled through variation of the frequency (and intensity) of light incident upon it. The rapid production of novel probes has quickly established a set of photoswitchable molecules that cover the whole visible spectrum.[76] The ability to express protein-based photoswitchable dyes has greatly helped in biological studies, such as monitoring the intercellular dynamics of proteins.[77]

6.1.4.1 Mechanisms of Photoswitching

Photoactivation (and thus photoswitching) can occur through various different mechanisms. For example, UV light can be used to break (photocleave) the bond between the spiro-carbon (sp^3) and the pyran (oxazine) moiety in spiropyrans (spirooxazines). This changes the orbital hybridisation of the carbon atom from sp^3 to

sp^2. In turn, this causes formation of a planar merocyanine molecule which has a conjugated π-system. When the UV light is removed the initial spiropyran or spirooxazine is thermally recovered. Figure 6.3 shows a schematic of this process, the reversible nature of which has led to proposals for these molecules to be used in applications such as memory and switches.[78]

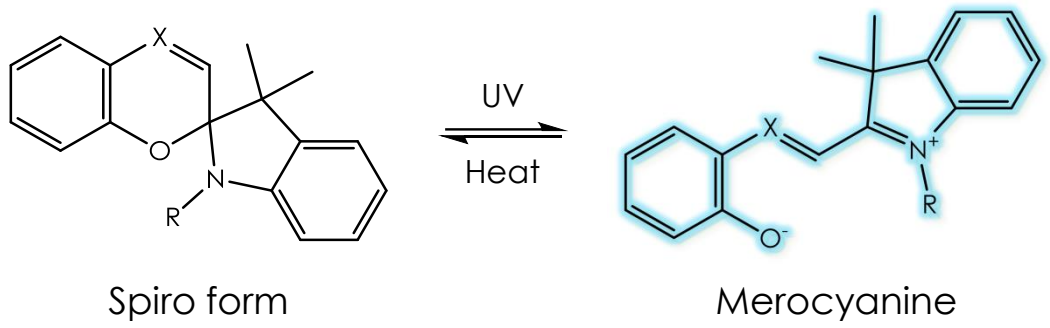

Figure 6.3 – UV irradiation of a colourless spiropyran (X = CH) or spirooxazine (X = N) causes photoconversion to a coloured merocyanine.[68,79]

Trans-cis isomerisation is known to be involved in the photoactivation mechanism of KFP-1.[74] Proton transfer has also been implicated as an important process involved in the photoswitching of fluorescent proteins.[80–82]

Recently, an electron paramagnetic resonance study has highlighted the importance of radical intermediates in the reversible photoswitching of rhodamine, oxazine, and thiazine dyes.[83] The fluorescent state of oxazine dyes has also been shown to be controllable with the addition of reductant or oxidant.[84] Other environmental factors, such as temperature or pH, may also influence photoswitching behaviour. The surrounding medium can therefore play an important role in the ability of a system to photoswitch. In some instances a switching buffer is required to ensure photoswitching is viable, as is the case with the Cy3-Cy5 pair used in this study. Details of the components of the Cy3-Cy5 switching buffer are discussed later in §6.2.5.

There are therefore potentially many elements which contribute to photoswitching efficiency in a particular system. This can make it difficult to pin-point the precise mechanism which causes photoswitching and, additionally, can greatly increase the difficulty of obtaining appropriate conditions within the system being studied. For instance, living cells require a particular environment to survive and so this precludes the use of extreme photoswitching conditions.

6.1.5 Photoswitching and Super-Resolution

The production of photoswitchable (photoactivating) molecules (or combinations of molecules) has been crucial for the development of reconstruction super-resolution techniques such as photoactivated localisation microscopy (PALM) and stochastic optical reconstruction microscopy (STORM). Photoswitching allows temporal discrimination of proximal fluorophores that would otherwise be spatially unresolvable. Precise localisation of individual fluorophores is therefore possible, even when label density is very high. This type of technique is sometimes referred to as single-molecule localisation microscopy (SMLM);[2] the general process of which is described below.

Figure 6.4 shows a schematic of the principle behind reconstruction super-resolution localisation techniques. Initially all fluorophores are in a dark state (not shown). Then a small sub-set of fluorophores is activated within a field-of-view at time t_1 (Figure 6.4a). Each active molecule creates a point-spread-function (PSF, see §6.1.3) in the image. Localisation of each molecule is achieved by fitting the PSF to a model function, such as a 2D Gaussian (Figure 6.4b).[85] Emitting fluorophores then return to the dark state. A new sub-set of fluorophores is activated within the same field-of-view at time t_2. Again, each active (emitting) molecule is localised. All localised molecules are accumulated together in a single image (Figure 6.4c). This process is repeated many times ($n = 10^2$-10^4) and a full picture of the system is gradually built-up. Resolution in the final (reconstructed) image is far superior to the diffraction limited image obtained *via* simultaneous excitation of the whole population of molecules. Localisation accuracy using these techniques is limited by the number of photons which are detected within each pixel but is typically of around 20-50 nm.[86] In accordance with Nyquist-Shannon sampling theory,[87] the density of fluorophores should be high to obtain the greatest level of detail.[33] It is, however, important that the PSFs of active molecules do not interfere with each other, as seen at time t_6 in Figure 6.4a. In a real experiment these overlapping molecules would be unresolvable and so either the frame would have to be dismissed or a single localisation would occur at the wrong position. Some more advanced analysis methods can handle a limited number of overlapping PSFs but, even so, activation conditions generally have to be kept fairly weak. This can result in empty frames, as seen at time t_5 in Figure 6.4a, and can mean very long acquisition times are required to ensure that enough molecules are localised to obtain a continuous, or full, image.

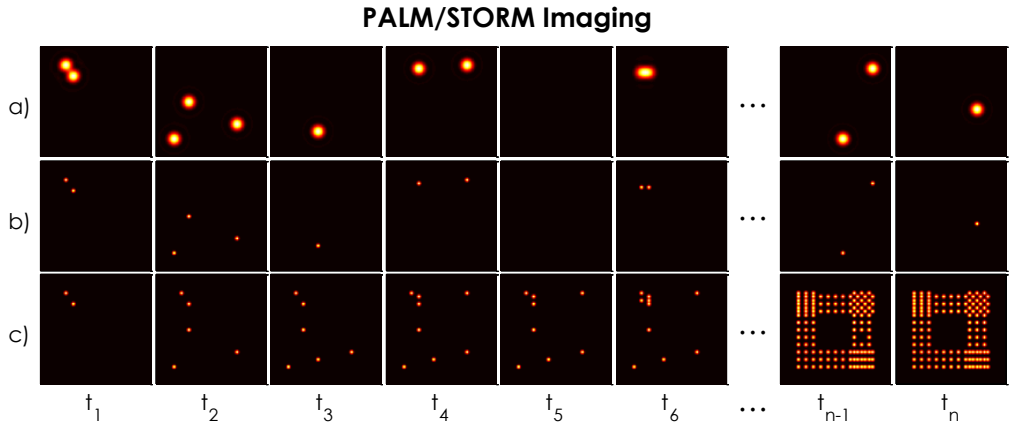

Figure 6.4 – Reconstruction super-resolution techniques such as PALM and STORM rely on stochastic localisation of isolated fluorophores. A total of n frames are recorded to produce a sequence of images. a) A sparse sub-set of the fluorophore population is photoactivated during each frame. b) At each frame time (t_i) molecules are identified and localised. c) The accumulated image of localised molecules has significantly higher resolution than would be possible with simultaneous excitation of all fluorophores; also see Figure 6.5. Note that all fluorophores are in their undetectable, dark state at times between the frames shown.

Figure 6.5 shows a simulation where fluorophores were either illuminated simultaneously or localised stochastically; the additional information that can be extracted is clear to see. In particular, the varied patterning in the corners of the square is completely lost in the conventional image. A movie of the simulation can be found in Supplementary Information (CD:\Movies\demoPALM\squarePALM).

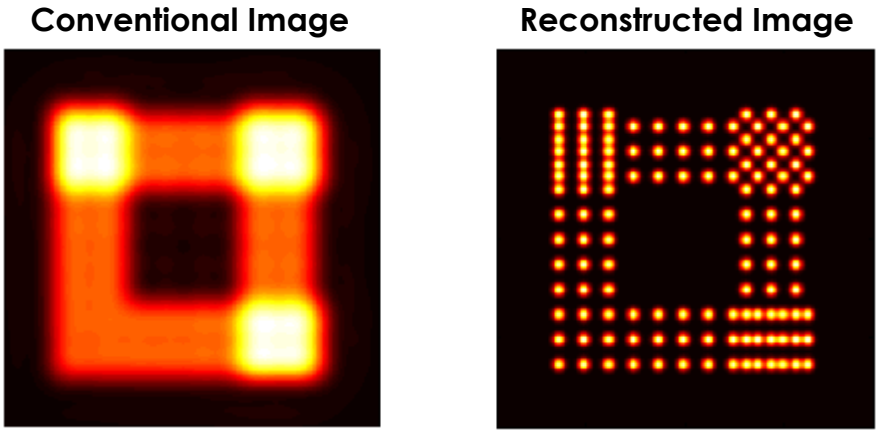

Figure 6.5 – Comparison of images produced by simultaneous excitation of all fluorophores (left) and by stochastic localisation (right). The conventional image lacks details which can be resolved in the reconstructed image.

6.1.6 Super-Resolution Imaging and DNA

Super-resolution microscopy offers the possibility of direct visualisation of the (quaternary) structure of DNA. This could have great impact in the use of DNA as a probe of the structural configuration of the numerous, and often important, systems in which it plays a crucial role.

Exogenous probes which utilise DNA as a scaffold have useful advantages, particularly for studies of biological systems. Firstly, the dimensions of DNA mean that it is an excellent system to use for super-resolution technologies; microscopic structures can be resolved with nanoscale detail. This would be particularly useful in probing structures which are fairly large but also highly complex, such as chromatin (the complex of DNA and proteins that constitute the nucleus of the cell). Secondly, DNA is as natural component and so there should be little disruption to a biological system once it has been introduced. Thirdly, the specificity of base-pairing means that DNA sequences can be designed to target a specific process within the system. For example, a recognition sequence could be utilised to study a particular DNA-enzyme interaction. Finally, DNA is involved in many important processes within cells and so being able to visualise where it is located and how it is being used could be very useful for improving our understanding of cellular mechanics.

There are, however, complications associated with the use of DNA. In particular, synthesising DNA (artificially) in excess of around 100 base pairs is difficult and this limits the length of substrates produced this way. Longer substrates are required to improve duplex stability (melting temperature) and also to allow better selectivity. The challenges of creating long DNA strands means that costs can escalate quickly, even before trying to incorporate potentially expensive labels at high density.

The persistence length (a measure of the step-size of the random walk of a polymer) of the DNA double helix is around 50 nm (~150 bp) and so short substrates will effectively assume a rigid, linear conformation. While this could be beneficial in some situations (such as measuring distances), this might hinder the use of DNA in obtaining structural information on smaller scales.

Super-resolution techniques require a high density of labelling but this is difficult to achieve in DNA because there are few efficient labelling methods available. Additionally, there is the potential structural disruption caused by having many dyes incorporated into (or near) the double helix and so an appropriate label must be used.

One strategy has been to label DNA-associated proteins to discern the topological features of the DNA under investigation.[17] This technique is, however, reliant on the fact that the protein used is present in a high concentration. Also, it may not even be possible to use such proteins in other systems.

Alternatively, it is possible to use intercalating or minor groove binding dyes which can have a significant increase in fluorescence upon binding to DNA.[8,9] The use of such dyes can also be problematic. For example, production of singlet oxygen by externally bound dyes can cause damage, such as single-strand breaks, to the DNA.[88] Perhaps more importantly, from an imaging perspective, is that the dyes may not be DNA-specific and so could also interact with RNA.[89] A lack of specificity means that the identity of the imaged structure is ambiguous.

A further approach is to use click-chemistry.[90] This method has been used to successfully image chromatin with super-resolution detail within a fixed cell but also has some limitations.[13] For example, incorporation of 5-ethynyl-2'-deoxyuridine (EdU) is random (with the exception that only thymidine sites should be labelled) and so the exact position of the fluorophore is not known. Additionally, the two-step nature of the click-chemistry process (first incorporating the modified base and then reacting ethynyl and azide groups) adds another level of stochastic uncertainty. The concentration ratio of EdU to azide-group labelled fluorophore is therefore very important. The ratio needs to be optimised to ensure labelling is dense enough to obtain continuous structures but is also sparse enough to allow localisation of individual fluorophores during acquisition. It is unclear how unincorporated EdU and unbound fluorophores are removed from the cell to ensure that fluorophores are only found on DNA sites and not found in isolation or as aggregates.

6.1.7 Contrast and Optical Lock-In Detection Microscopy

It is obviously important to be able to distinguish between signal and noise, the latter of which can, unfortunately, be generated in a number of ways. Shot noise (or photon noise) is caused by the stochastic nature of fluorescence emission which causes temporal fluctuations in signal intensity. Fluorescence from species outside of the focal plane (or not at their desired target) and autofluorescence of natural components (such as tryptophan, tyrosine, and phenylalanine residues) can result in very large background levels.[3] Noise can also be caused by light scattering within the sample or from imperfections in detection system, including chromatic aberration from the optics or electrical fluctuations in the CCD camera.

Optical lock-in detection (OLID) microscopy is a recently developed technique that greatly enhances contrast within an image by isolating signal from noise.[3] The premise of OLID is remarkably simple. Each individual pixel is compared to a reference waveform. Pixels that show a high (low) correlation to the reference are given a high (low) value. Once all pixels have been evaluated, a final correlation image can then be created. Figure 6.6 shows a schematic of this process.

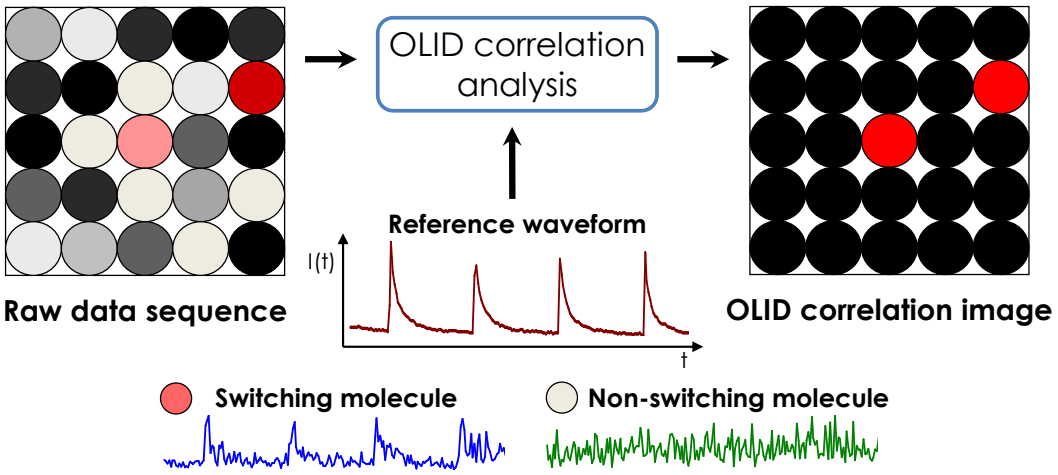

Figure 6.6 – Schematic diagram of the principle behind optical lock-in detection (OLID) microscopy. Each pixel in a raw data sequence is compared to a reference waveform. Only correlated, switching molecules are present in the final correlation image. The signal-to-noise ratio is greatly increased. Traces show the temporal change in intensity of the reference waveform (brown), a switching molecule (blue) and a non-switching molecule (green).

The correlation coefficient, $\rho(x,y)$, for each pixel is calculated using Equation 6.2;[3]

$$\rho(x,y) = \sum_{t=1}^{N} \frac{\{I(x,y,t) - \mu_{I(x,y)}\}\{R(t) - \mu_R\}}{\sigma_{I(x,y)} \cdot \sigma_R} \qquad 6.2$$

where $I(x,y,t)$ is the intensity of the pixel at position (x,y) at time t; $R(t)$ is the intensity of the reference waveform at time t; N is the total number of frames in a sequence; $\mu_{I(x,y)}$ and μ_R are mean values of the individual pixel and reference waveform, respectively; and $\sigma_{I(x,y)}$ and σ_R are standard deviation values of the individual pixel and reference waveform, respectively. Correlation coefficients can range from a value of 0 to 1. A more mathematical description of the analysis used for OLID microscopy is presented in Appendix 6, §6.1.

As correlation is paramount to the success of OLID, it is necessary to use a sample which can be efficiently controlled during the course of a sequence. Photoswitchable molecules are ideally suited to this task because they undergo a large, and thus detectable, change at a well-defined moment. All photoswitchable molecules in a sample will show synchronised behaviour and so they can all be correlated to the same reference waveform.

The reference waveform can be created in a number of different ways. One method is to use the temporal intensity trace of a location in the sequence which is known to have a high concentration of photoswitchable molecules in it. If necessary, this could be some introduced probe, such as a densely labelled bead.[3]

It is important to realise that the final correlation value of each pixel is independent of the intensity range which it sampled during the sequence. Thus, large background noise, which randomly fluctuates, is greatly diminished in the final correlation image. In contrast, a weak signal, which is highly correlated to the reference waveform, is greatly enhanced by the OLID analysis. OLID has already been used to produce high contrast images in systems such as cells[91] and even living Zebrafish.[3] The technique has great promise for future studies where probe numbers are low (or their intensity is weak) and background signal is high.

It should be noted that OLID microscopy is not a super-resolution technique. The highest achievable resolution is determined by the effective pixel size of the detector. Even this level of resolution is unlikely to be able to be achieved because the PSF of the microscope will increase the correlation values of neighbouring pixels.

Although the outcome is somewhat different to super-resolution techniques, OLID microscopy will also benefit greatly from the development of molecules with improved photoswitching capabilities.

6.1.8 CyDNA

While current solid state synthesis methods are limited to producing only short oligonucleotides (100-200 bases) with high fidelity, polymerase chain reaction (PCR) can be used to replicate long DNA sequences. In addition, the use of a certain DNA polymerases can allow the incorporation of base analogue dyes to create a labelled DNA sequence.[92–95] Ramsay *et al.*[16] recently introduced a polymerase, E10, which could produce CyDNA; a substrate where many natural cytosine bases (dC) are replaced by a cyanine dye base analogue, either Cy3-dC or Cy5-dC (Figure 6.7). Substrates could be constructed up to 1.3 kbp. Evolution of the polymerase was achieved using a novel

technique known as short-patch compartmentalised self-replication (spCSR). To improve accommodation of the bulky dye groups attached to the nucleotides the polymerase active site was targeted during evolution. As it was predicted that there would be the creation of some specialised polymerases (ones that could incorporate both Cy3- and Cy5-type labelled nucleotides) potential clones were screened. Fidelity of the polymerase was shown to diminish when additives used during spCSR (formamide, glycerol, DNase, and DTT) were absent. CyDNA structure was investigated using restriction endonucleases. DdeI (which cuts at the prime in the sequence C'TNAG) failed to cut but this was thought to be due to the bulky dye group rather than structural abnormality. MseI (T'TAA) cut efficiently suggesting that at least AT structure was of typical double-helical form. The melting temperature of CyDNA was found to be close to that of natural DNA suggesting that double-helical structure was maintained even in the presence of many dyes. For clarity, CyDNA substrates containing only Cy3 or Cy5 dyes were given the notation Cy3DNA and Cy5DNA, respectively.

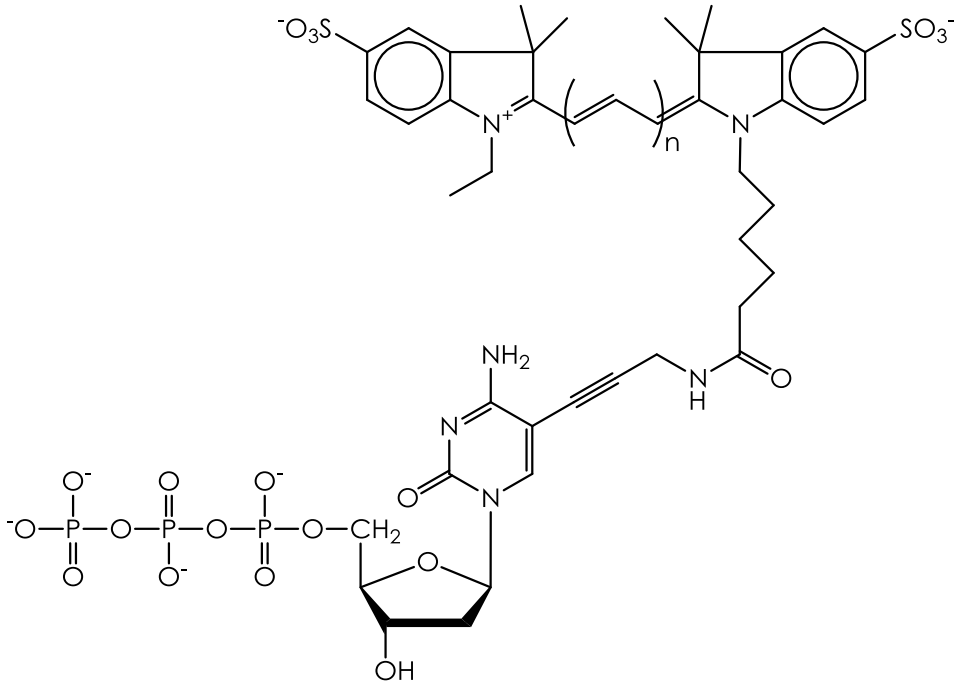

Figure 6.7 – General structure of 5-amino-propargyl-2'-deoxycytidine 5'-triphosphate coupled to a cyanine dye; Cy3-AP3-dCTP (n = 1, Cy3-dCTP) and Cy5-AP3-dCTP (n = 2, Cy5-dCTP).

One of the most desirable properties of CyDNA is that fact that it has the potential to meet the requirements for super-resolution imaging of DNA. Fluorophores are densely labelled and, in the presence of an oxygen scavenging system and source of a

primary thiol, proximal (within 3 nm) Cy3 and Cy5 dyes have been shown to be a highly efficient photoswitch.[19] The mechanism of photoswitching is, however, not well-established. The following two sections discuss what is currently known.

6.1.8.1 Cy5 Dark State

Mass spectroscopy was used to show that a dark state of the Cy5 molecule was produced *via* attachment of a primary thiol, such as cysteamine, to the polymethine bridge.[96] Attachment occurred to the second (or fourth) carbon in the bridge through an excited state of the Cy5 molecule. Corresponding to the formation of the dark state, there is an increase in absorbance at around 310 nm and reduction at around 650 nm.[96,97] It is worth noting that the shorter polymethine bridge in Cy3 precludes the formation of a dark state in this molecule.

Bates, Blosser, and Zhuang[19] showed that the rate of dark state production was linearly dependent on potassium iodide concentration. This suggests that a triplet state was involved in the transition, since iodide strongly enhances intersystem crossing *via* heavy-atom-induced spin-orbit coupling.[98,99] An oxygen scavenging system is often used in studies involving Cy5 to reduce photobleaching effects.[96,100] Additionally, the removal of oxygen would likely increase the lifetime of a triplet state species. It has been known for over half a century that molecular oxygen can quench excited states.[101,102] Removal of oxygen should, therefore, also increase dark state production.

Bates, Blosser, and Zhuang[19] also found that the presence of isoascorbate, a radical quencher, dramatically reduced the rate of switching to the Cy5 dark state. This observation would suggest that, in addition to a triplet state, radicals are also important intermediates in the photoswitching mechanism of Cy5. The requirement for radical (and triplet) formation to produce the Cy5 dark state has been further corroborated by the work of Van de Linde *et al*. which provides evidence that similar intermediates are involved in the photoswitching mechanism of oxazine, thiazine, and rhodamine dyes.[83]

It has been suggested that *trans-cis* isomerisation produces a Cy5 dim state (that is, a state where fluorescence intensity is partially diminished).[100] This state has been observed in other studies.[103,104] There is also evidence of Cy3 undergoing a similar isomerisation from the first excited state.[105] Formation of the Cy5 dark state was, however, found to be independent of solvent viscosity.[19] Increasing the viscosity of the solvent should hinder molecular motion and therefore reduce the rate of isomerisation. *Trans-cis* isomerisation is perhaps an off-pathway transition that competes with production of the dark state.

6.1.8.2 Cy3-Cy5 Photoswitching

Recovery from the Cy5 dark state has been shown to occur *via* direct excitation at 337 nm, 488 nm, and 532 nm[97] and also by using a filtered UV lamp (325 ± 30 nm).[96] It is therefore possible for isolated Cy5 to reversibly transition between a dark state and a fluorescent state. There is, however, no control over when transitions occur and so Cy5 molecules randomly 'blink' on and off.

The presence of a Cy3 molecule can, under certain conditions, significantly improve Cy5 switching efficiency. In addition to a source of a primary thiol, Cy3 and Cy5 dyes must be in close proximity to each other and an oxygen scavenging system must be present.[19,97] The power of the photoactivating laser can be reduced several orders of magnitude when switching *via* excitation of the Cy3 molecule.[97] Photoactivation occurs only when Cy3 molecules are being excited and so the process is controllable.

While Cy3 and Cy5 are known to undergo Förster resonance energy transfer (FRET),[106] it seems unlikely that it is involved in the main pathway for recovery from the dark state. Transition to the dark state will significantly diminish the spectral overlap of Cy3 emission and Cy5 excitation and thus reduce FRET efficiency. Additionally, direct excitation of Cy5 should not be so inefficient in recovering the fluorescent state if FRET is really a major pathway. This evaluation is consistent with the fact that switching efficiency falls off much faster with increasing dye separation than theoretical FRET efficiency for the pairing.[19] The quicker drop-off rate does, however, mean that it is possible to measure distances smaller than a conventional FRET ruler would allow.

It is clear that further investigation is required to pin-down the complete switching mechanism of the Cy3-Cy5 pair.

6.1.9 CyDNA and OLID

For OLID correlation analysis to be effective, it must be possible to detect signal at the single-pixel level. Noise levels are relatively high within a single pixel and so the probe used for these studies must be highly fluorescent. The Cy3-Cy5 pairing had already been used in single-molecule studies and, thus, had been shown to give a strong fluorescence signal.[17–19] Photoswitchable CyDNA (psCyDNA, see §6.2.2) combines this property with a high density of labelling. In addition to increasing CyDNA brightness (despite possible quenching problems, which are discussed later in §6.3.2), the high density of fluorophores should mean that random switching events do not significantly

alter fluorescence intensity. In contrast, photoactivated switching will produce a large change in intensity and so psCyDNA has great promise as an OLID system.

It is worth noting that conditions which are optimal for OLID imaging are somewhat contrary to those of super-resolution imaging, where it is important to have sparse activation events. The apparent conflict of the two methods can be resolved for CyDNA by changing the activation conditions (for example, changing laser intensities or buffer concentrations) or by altering the dye distribution (for example, altering Cy3 dye numbers to change the number of simultaneous activation events).

The next section will describe the procedures and apparatus used to develop and assess reversible fluorescence photoswitching in CyDNA.

6.2 Materials and Methods

In all cases water was of at least of HPLC grade. Chemicals were either from Sigma-Aldrich or Fisher Scientific (unless otherwise stated) and used as received. Cy3-dCTP, Cy5-dCTP, and CyDNA were donated by Dr. Philipp Holliger[107] or synthesised as described below.

6.2.1 Synthesis of CyDNA

6.2.1.1 Polymerase Chain Reaction (PCR)

A mutant *Pyrococcus furiousus* (*Pfu*) polymerase, E10, was used for polymerase chain reaction (PCR).[16] Samples were prepared in ThermoPol buffer (New England BioLabs) and 1% (v/v) formamide. Plasmid (10 ng or 1 µg) from *Thermococcus gorgonarius* (*Tgo*), pASK*Tgo*T, was used as template. Primers (0.4 µg) used were 5'-ACC ACC GAA CTG CGG GTG ACG CCA AG-3' and 5'-GCT GAG GAA GGC CTA CGA GAG-3'. Water was used to give a total reaction volume of 100 µl. The concentration of Cy5-dCTP and dNTPs (deoxynucleotide triphosphate; d(A, C, G, T)TPs) in the reaction mixture was varied to give different label densities. Table 6.2 summarises the concentrations used in this study.

Sample	[Cy5-dCTP] /µM	[dCTP] /µM	[d(A, T, G)TPs] /µM
1	1	9	10
2	2	8	10
3	5	5	10
4	25	25	50

Table 6.2 – Polymerase chain reaction sample composition.

It is worth noting that Sample 3 and Sample 4 result from PCR mixtures with the same ratio of labelled to unlabelled dCTP; however, the incorporation of Cy5-dCTP into CyDNA (and thus the labelling density) is higher in Sample 4 (as can be seen later in Figure 6.12). The exact cause of this effect is unknown but it suggests that the relationship between the incorporation efficiency of the enzyme and the nucleotide concentration in the PCR mixture is not the same for dCTP and Cy5-dCTP. A higher incorporation efficiency of base analogues compared to natural bases has been observed before in other PCR mediated synthesies.[95,108] It is possible that dye aggregation may play some role in this effect.[109]

Cy3DNA was prepared using the same conditions as Sample 4 with the exception that Cy5-dCTP was replaced with Cy3-dCTP.

PCR was performed in a thermal cycler (Eppendorf, Mastercycler). Reaction conditions were as follows: 94°C for two minutes (min), then 50 cycles of {94°C for 10 seconds (s), 55°C for one min, and 72°C for 20 min}. Samples were held at 4°C until further use or stored at −20°C. A very brief explanation of the stages involved in the PCR process can be found in Appendix 6, §6.6.

The amplified fragment had a GC content of 50% and so a maximum of 25% labelled base pairs could be achieved using the above conditions.

PCR products were purified by ethanol precipitation with 2.5× volume ethanol and 0.1× volume sodium acetate (pH 5.2). Samples were re-dissolved in Tris-EDTA buffer and further purified with an S-400 column (GE Healthcare).

6.2.1.2 Agarose Gel Electrophoresis

PCR product size was analysed using agarose gel electrophoresis. Gels were prepared using 1% (w/v) agarose in Tris/Borate/EDTA (TBE) buffer. Sample volume was either 12 µl or 20 µl. Gels were run for around 30 min in constant voltage mode (100 V, 108 mA). Ladders (Fermentas, SM0193 and New England BioLabs, QuickLoad 1 kb DNA) were used to size the PCR products. SYBR® Safe dye (Invitrogen) was used to visualise bands.

6.2.2 Photoswitchable CyDNA (psCyDNA)

Two different strategies were used to produce photoswitchable CyDNA (psCyDNA). The first involved hybridising Cy3DNA and Cy5DNA strands while the other method simply used a mixture of Cy3- and Cy5-labelled dCTP during PCR.

6.2.2.1 Cy3DNA and Cy5DNA Hybridisation (Cy3/5DNA)

Equal molarities of Cy3DNA and Cy5DNA were mixed together in solution. The duplexes were then melted at 95°C for one min and then cooled to room temperature at 1°C per min in a thermal cycler (Eppendorf, Mastercycler). The resulting sample contained Cy3DNA, Cy5DNA, and Cy3/5DNA (a heteroduplex containing strands of Cy3DNA and Cy5DNA). Assuming that hybridisation efficiency was independent of dye type, permutations of single strand pairings were expected to produce Cy3/5DNA in twice the concentration of Cy3DNA and Cy5DNA.[i]

[i] 2A:A' + 2B:B' → A:A' + B:B' + (A:B' + B:A'); X' complementary strand to X

6.2.2.2 *Cy*DNA*

Using the same PCR conditions described above (§6.2.1.1), Cy*DNA was synthesised by using the following dNTP mixture: 10 μM d(A,T,G)TP, 5 μM dCTP, 2.5 μM of Cy3-dCTP, and 2.5 μM of Cy5-dCTP. This produced CyDNA which had a random incorporation of Cy3 and Cy5 dyes in both strands.

6.2.3 Preparation of CyDNA-Coated Beads

Silica beads 1.2 μm (or 1.5 μm) were added to poly-L-lysine solution (200 μl). Beads were then centrifuged for 10 s in a (micro) fixed-angle centrifuge. Supernatant was removed before refilling to around 200-500 μl with phosphate buffered saline (PBS). Three or four centrifuge cycles were performed. 1 μl (or 9 μl) of Cy5-dCTP (CyDNA) solution was used to coat the beads. 15 μl of bead solution was added to a glass coverslip for experiments.

6.2.4 Photophysical Studies

Absorption measurements were recorded on a UV-Vis spectrometer (Cary 50 Bio, Varian). A matched cuvette filled with PBS was used as a blank for measurements. Absorption measurements for Cy3 (Cy5) samples were made between 400 nm (500 nm) and 700 nm (800 nm). Scan speed was 600 nm per minute.

A sample of Cy3-dCTP (Cy5-dCTP) was diluted with PBS to create a point of equal extinction with a Cy3DNA (Cy5DNA) sample. Emission spectra were recorded by exciting at the wavelength of the point of equal extinction (532 nm for Cy3 and 644 nm for Cy5).

Excitation and emission spectra were recorded on a FluoroMax spectrophotometer (Horiba Jobin Yvon). Excitation and emission slits were set to 5 nm. Integration time was 0.2 s and step size was 1 nm. Spectra were collected over various ranges dependent on sample. Spectra were corrected using instrument calibration files.

6.2.5 Microscopy

Switching buffer was prepared by dissolving glucose (10% w/v), glucose oxidase (0.5 mg ml^{-1}), catalase (40 μg ml^{-1}), and cysteamine (3.8 mg ml^{-1}) in PBS (pH 7.4). Glucose oxidase, glucose, and catalase acted as an oxygen scavenging system while cysteamine was used as a source of a primary thiol (see also section §6.1.8.2).

Sample molecules (20 µl, ~0.4 ng µl^{-1}) were immobilised on glass coverslips (#1 or #1½) coated in poly-L-lysine. CoverWell™ imaging chambers (Grace Bio-Labs) containing switching buffer were sealed using the sample coverslip.

A schematic of the experimental setup is shown in Figure 6.8. All measurements were performed on a Nikon TE2000 Eclipse inverted microscope. Imaging was achieved with a 633 nm laser (Helium-Neon, Coherent, #31-2140-000, 18 mW) while a 532 nm laser (Samba, Cobolt, 0.1-0.3 mW) was used for photoactivation. A dichroic mirror (545DM) was used to co-align the two continuous wave laser beams used during experiments.

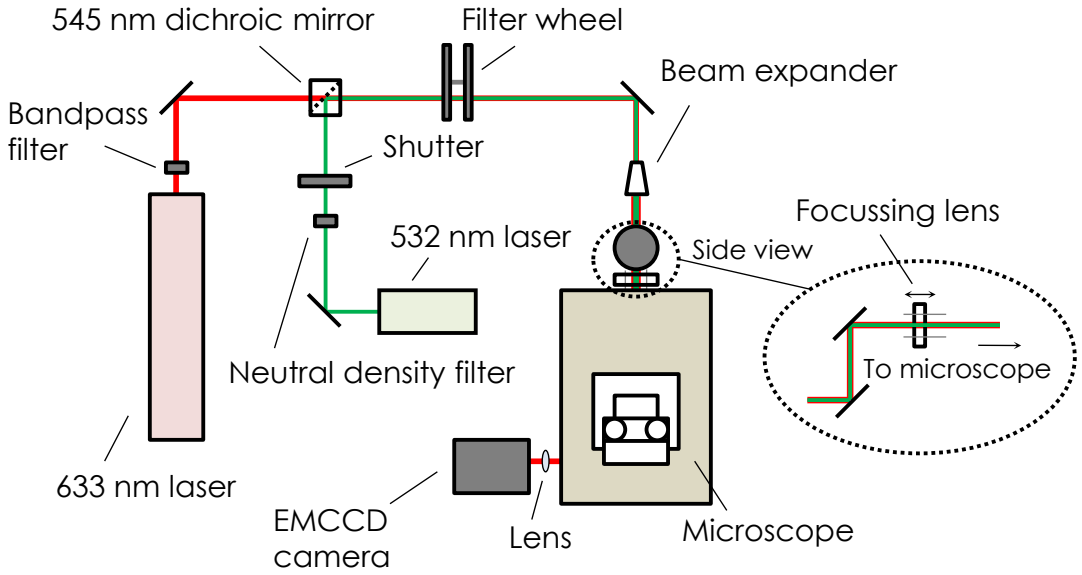

Figure 6.8 – Schematic of the microscopy set-up.

The objective lens (Nikon, CFI Apochromat TIRF, 60x/1.49 Oil) was corrected for coverslip thickness. Excitation light was filtered out using a dichroic mirror (633DM). Filtering was improved by using an additional long pass filter (645LP).

Detection was achieved using an electron-multiplying charge-coupled device (EMCCD) camera (Andor Luca (S)) which allowed wide-field images (658 × 496 pixels) to be collected. Total magnification was around 198, relating to an individual pixel size of around 74 nm × 74 nm and a wide-field image size of 49 µm × 37 µm. Detector pixel size is important for super-resolution imaging because the pixel has to be large enough to collect enough photons while being small enough to achieve the best possible resolution. Optimal pixel size for localisation is typically considered to be the standard deviation of the PSF of the system.[85]

In TIRF mode, images were recorded using *Andor SOLIS* software.[110] Typically each sequence was recorded over 200 frames. Exposure time was 100 ms. During experiments a mechanical shutter was used to control the pulses of the 523 nm laser. Repetition rate and the duty cycle (the percentage of a cycle in which the shutter is open) could be adjusted for each experiment using an in-house LabVIEW[111] program. Movies of the raw data sequences can be found in Supplementary Information (CD:\Movies\dataSequences\).

The (mean) intensity within a region-of-interest (typically 5 × 5 pixels) or a single pixel during a sequence was obtained directly from *Andor SOLIS* to give single-molecule or single-pixel intensity traces, respectively.

A proof-of-principle super-resolution experiment was performed on a Cy3/5DNA molecule spin coated onto a poly-L-lysine coated coverslip. Dr. Cristina Flors collected and analysed the data for this experiment.

Detection and localisation of molecules for all microscopy experiments was achieved using IGOR Pro.[112] Molecules were identified and mapped using *Localizer* analysis software provided by Dr. Peter Dedecker.[113] Histograms of active molecules and molecule intensities were prepared in Microsoft Excel (2010).[114]

OLID images were prepared *via* a Java[115] program and then rendered using ImageJ.[116–118] The OLID analysis code was also ported to MATLAB.[119] This allowed the use of parallel computing during the analysis and so reduced overall computation time.

The results obtained from the experiments performed in this investigation will be presented in the next section along with a discussion their outcomes.

6.3 Results and Discussion

6.3.1 Confirmation of PCR Product by Agarose Gel Electrophoresis

Figure 6.9 shows the gel obtained for various CyDNA substrates synthesised using different dNTP concentrations.[j] Figures 6.9a and 6.9b show clear bands at the expected PCR product length of around 1.3 kbp. Figure 6.9c shows a smear rather than a band. No features were observed when using 25 µM Cy5-dCTP and 50 µM dNTPs, however, it was possible to extract DNA from this sample.[k] A possible explanation of the lack of a band in the gel is that the high density of labelling hindered the intercalating SYBR® Safe dye from binding to the DNA. Ideally, CyDNA would have been directly visualised without an additional dye – in accordance with Ramsay *et al.*[16] Unfortunately, this was not possible with the very low sample concentrations that were prepared during this work.

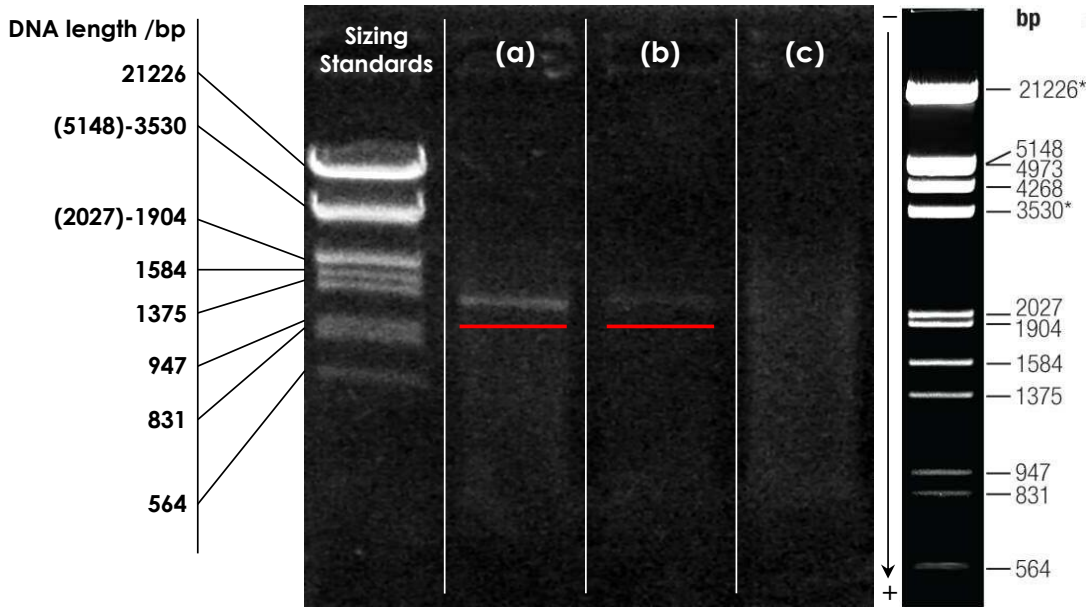

Figure 6.9 – Gel showing the result of PCR with various reaction compositions: (a) 1 µM Cy5-dCTP, 10 µM dNTPs; (b) 5 µM Cy5-dCTP, 10 µM dNTPs, and (c) 5 µM Cy5-dCTP, 50 µM dNTPs. The scale of the reference and experimental sizing standard bands is different. For clarity the bands in the gel have been highlighted with a red line.

---

[j] Results from initial PCR experiments suggested that the optimal conditions for producing CyDNA require the use of 1 µg of template and 10 µM of dNTPs. When using only 10 ng of template no product bands were observed.

[k] Absorption measurements of this CyDNA were consistent with CyDNA of the same estimated labelling density which had been supplied by Dr. Philipp Holliger.[107]

6.3.2 Photophysics of CyDNA

6.3.2.1 Formation of H-Aggregates

Figure 6.10 and Figure 6.11 show normalised spectra for samples of Cy3DNA and Cy5DNA at similar label densities, respectively.

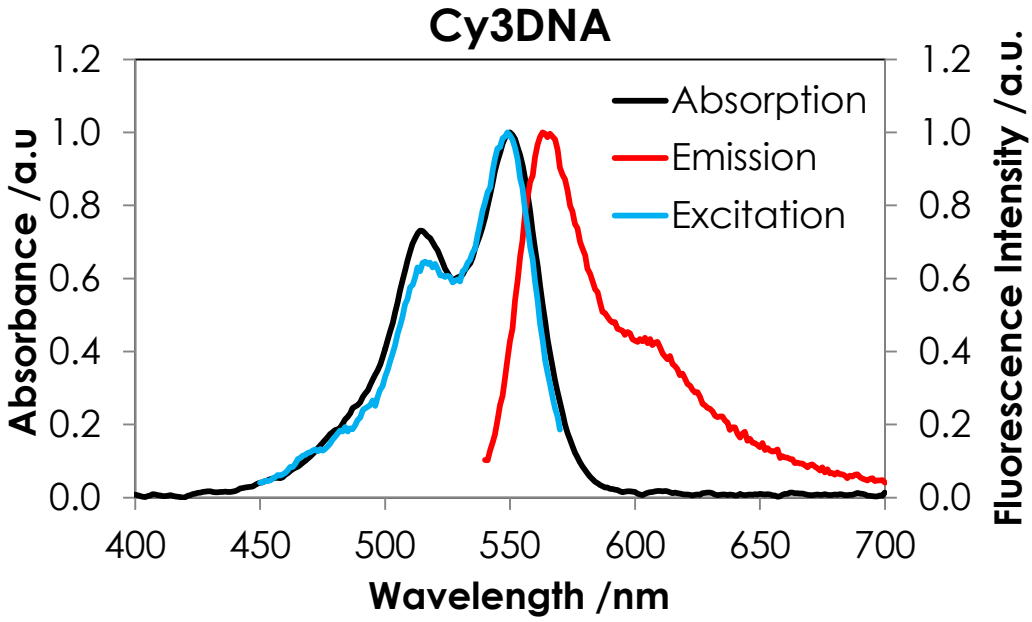

Figure 6.10 – Comparison of normalised absorption, excitation, and emission spectra for Cy3DNA.

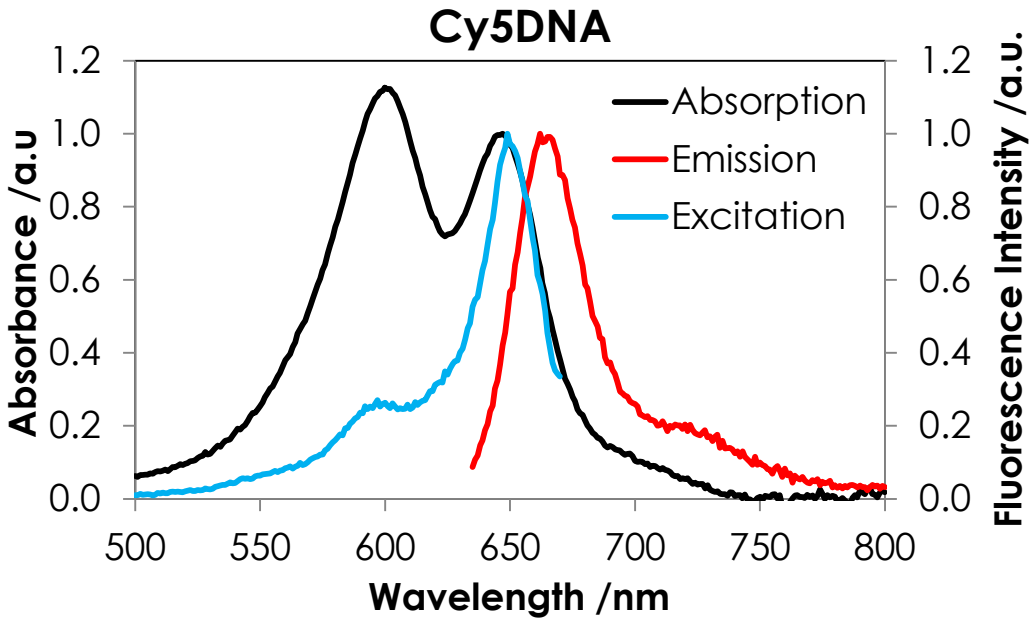

Figure 6.11 – Comparison of normalised absorption, excitation, and emission spectra for Cy5DNA.

Absorption spectra of CyDNA show an increase in the intensity of the shorter wavelength peak compared to excitation spectra which is not observed for Cy3-dCTP and Cy5-dCTP samples (Cy3-dCTP and Cy5-dCTP spectra are given in Appendix 6, §6.2). This increase has been attributed to the formation of non-fluorescent H-aggregates.[109,120–123] A brief explanation of H-aggregate formation is given in Appendix 6, §6.3. Consistent with the findings of Gruber *et al.*,[109] the much larger increase in the shorter wavelength peak shows that H-aggregates are more prevalent in Cy5DNA than Cy3DNA. Figure 6.12 shows that the 600 nm : 650 nm peak ratio increases with label concentration. This demonstrates that there is a concomitant increase in H-aggregate formation with label density.

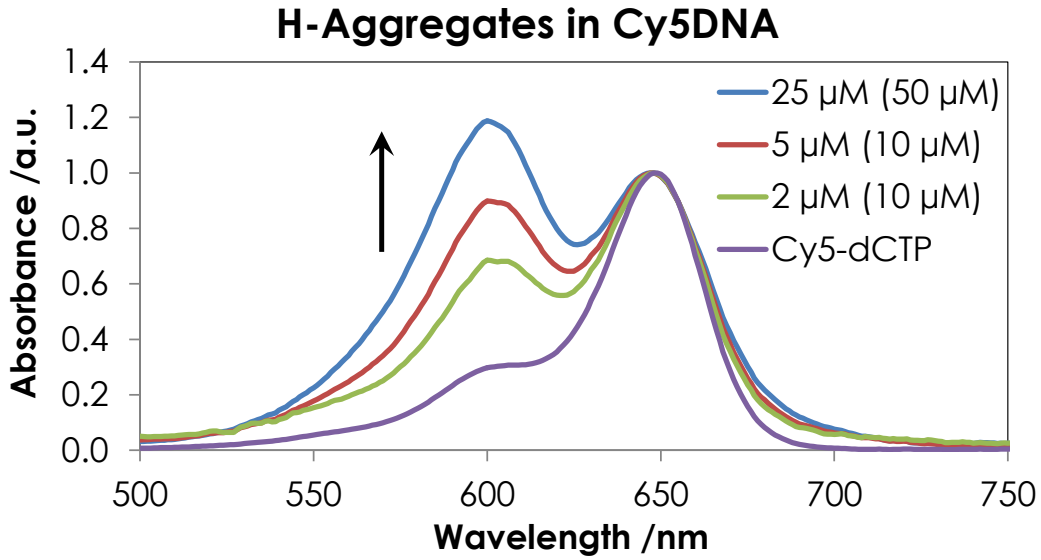

Figure 6.12 – Absorption spectra of Cy5-dCTP and Cy5DNA with various labelling densities. Spectra have been normalised to the $S_1 \leftarrow S_0$ peak maximum at around 650 nm. The legend shows the Cy5-dCTP (dNTP) concentration used during PCR. The vertical arrow indicates increasing label density.

6.3.2.2 *Additional Fluorescence Quenching Mechanisms*

Emission spectra for absorption-matched (that is, equally absorbing) samples of Cy3-dCTP (Cy5-dCTP) and Cy3DNA (Cy5DNA) are shown in Figure 6.13. Associated intensity maxima are given in Table 6.3.

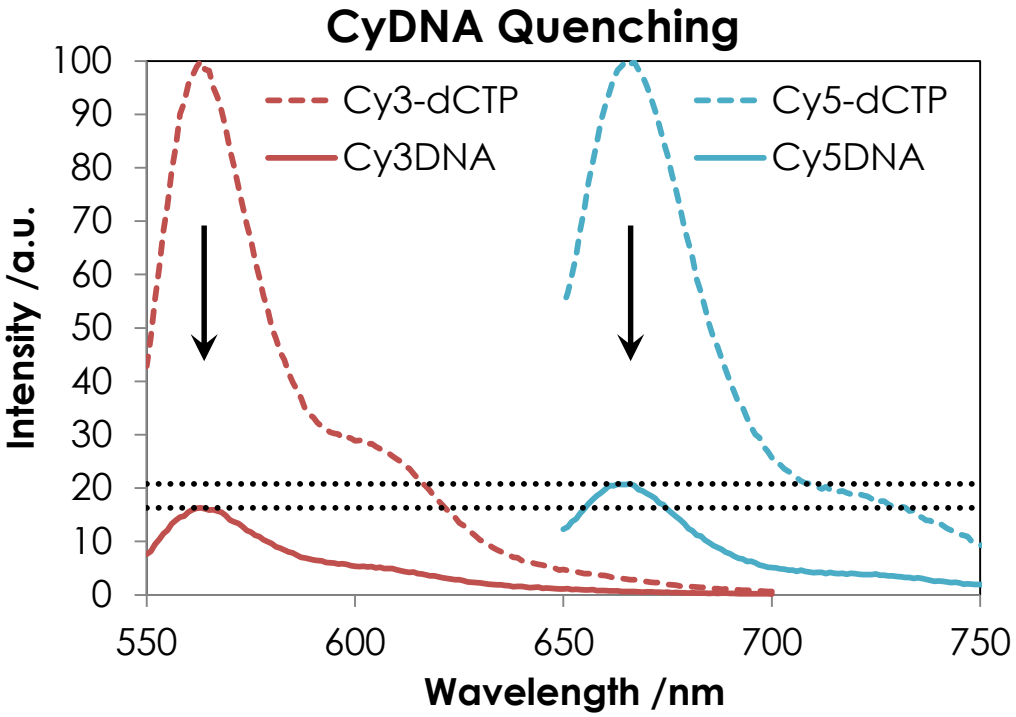

Figure 6.13 – Emission spectra of absorption matched samples were recorded by exciting at the wavelength of the point of equal extinction (532nm for Cy3 and 644nm for Cy5). Arrows indicate quenching of fluorescence in CyDNA compared to Cy-dCTPs. The black dotted lines are shown to aid in the comparison of the relative quenching in Cy3 and Cy5 species.

Dye	Cy-dCTP Maximum × 10^{-6} /cps	CyDNA Maximum × 10^{-6} /cps	Quenching /%
Cy3	8.14	1.33	84
Cy5	33.45	6.96	79

Table 6.3 – Comparison of emission intensity maxima obtained when exciting species at an isosbestic point (532nm for Cy3 and 644nm for Cy5). Due to a small residual absorption by H-aggregates at the excitation wavelength, the quenching percentages reported will be slightly overestimated.

For both Cy3- and Cy5-type samples, there is a large decrease in the maximum intensity observed for the CyDNA compared to the Cy-dCTP. This suggests both Cy3 and Cy5 dyes are significantly quenched when incorporated into the DNA substrate (at high density). As the excitation wavelength was outside of the H-aggregate absorption band this quenching must have been occuring through some other channel. Resonance energy transfer (RET) to a low-energy dark state could be one possible mechanism for

the observed quenching. Low-energy sites have previously been evident in low-temperature studies of conjugated polymers.[124] Despite containing a large number of absorbing species, efficient energy-transfer meant that the fluorescence behaviour of these systems was dominated by only a few low-energy chromophores. It is worth noting that dyes within CyDNA could interact with each other (and with natural bases) along the same strand or between strands; this perhaps helps to explain why quenching is so significant.

6.3.2.3 *Single-Molecule Intensity Traces*

Figure 6.14 shows examples of single-molecule intensity traces collected for Cy5-dCTP and Cy5DNA molecules. Each trace represents the temporal change in the average intensity of a localised region-of-interest (5 × 5 pixels) which was centred on a single molecule. Traces have been corrected for a background intensity of around 700 counts (per 100 ms) and show random intensity fluctuations as molecules switch between fluorescent and dark states. Additional single-molecule intensity traces can be found in Appendix 6, §6.4.

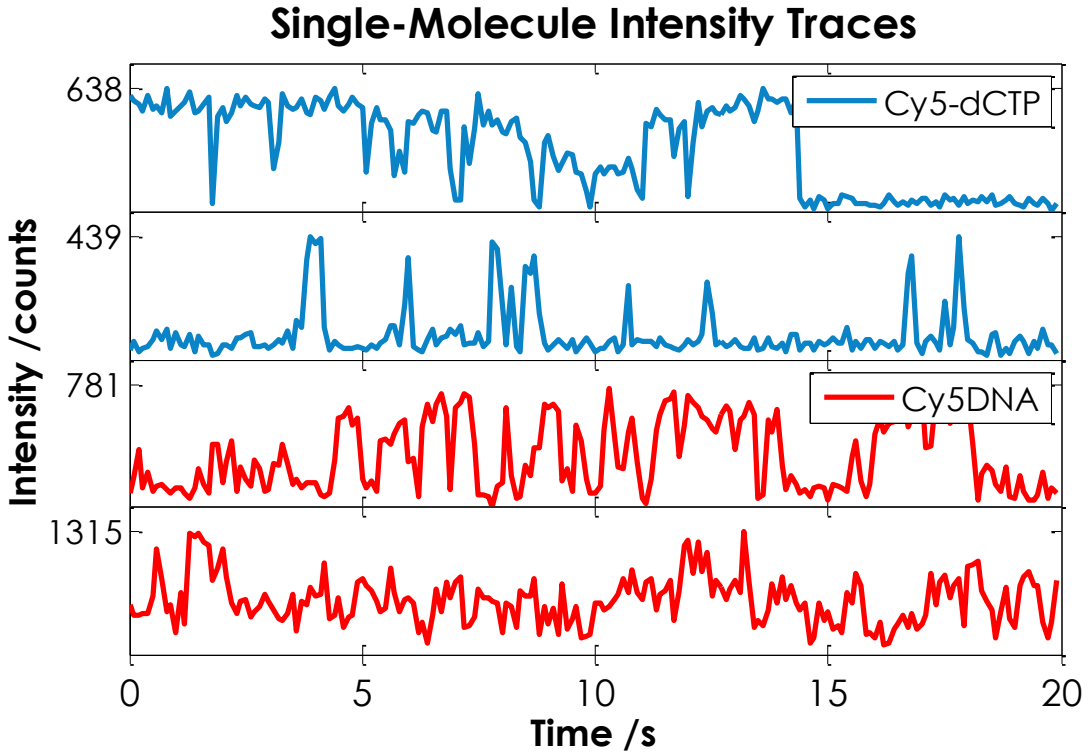

Figure 6.14 – Single-molecule intensity traces of Cy5-dCTP and Cy5DNA molecules under continuous 633 nm excitation. Traces are shown for two of each type of molecule. Traces were created by averaging a 5 × 5 pixel region-of-interest and have been background corrected. The y-axis values represent the maximum intensities for each trace.

Cy5-dCTP molecules generally show only two intensity levels suggesting that a single fluorophore is responsible for emission. In contrast, Cy5DNA molecules appear to have more variable intensity suggesting the presence of multiple emitters. The maximum intensity observed for Cy5DNA molecules is only slightly higher than that of the Cy5-dCTP molecules. This suggests that only one or two Cy5 dyes are fluorescent at any given time in Cy5DNA. As Cy5DNA is expected to contain hundreds of labels, this is further evidence of efficient quenching within this molecule.

6.3.3 Photoswitching of CyDNA: Ensemble and Single-Molecule Measurements

While the photoswitching mechanism of the Cy3-Cy5 pairing might still be unclear, the effect is real and measurable. Figure 6.15 shows that Cy3/5DNA and Cy*DNA can be efficiently photoswitched many times over the course of a 20 s sequence.

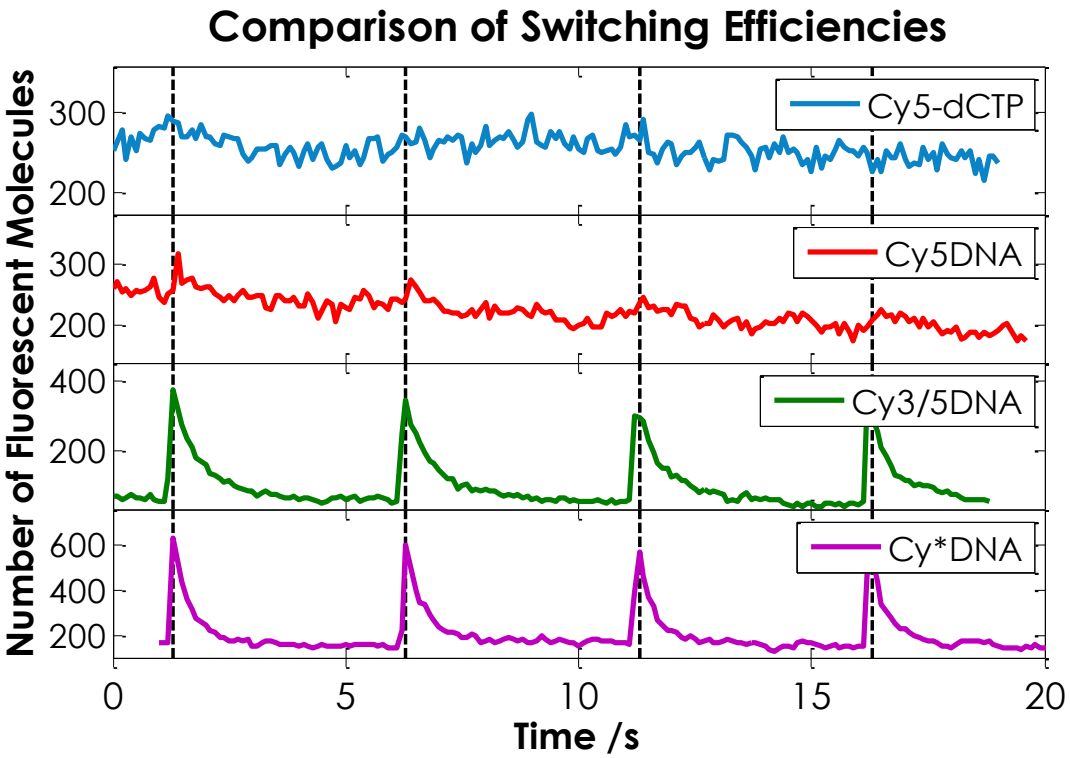

Figure 6.15 – A comparison of the photoswitching ability of Cy5-dCTP, Cy5DNA, Cy3/5DNA, and Cy*DNA molecules. Excitation at 633 nm was continuous while 532 nm light was pulsed every 5 s (dashed lines).

During experiments a 633 nm excitation source was continuously incident (this meant that Cy5 molecules equilibrated between dark and fluorescent states) while

532 nm excitation was pulsed every 5 s. The number of fluorescent molecules randomly fluctuates for a sample only containing Cy5-dCTP and there is little evidence of activation under the conditions used. For Cy5DNA there is a slight increase in the number of fluorescent molecules during the 532 nm pulses. Given that direct photoactivation has previously been shown to occur for Cy5, this effect is not completely surprising;[96,97] however, this could potentially disrupt OLID analysis as 'non-switching' molecules might be photoactivated as well as switching molecules. Importantly, the switching behaviour in Cy5DNA is weak and inconsistent (that is, it does not occur with each and every 532 nm pulse). This means that these random photoactivation events in Cy5DNA are insignificant when analysing the data over the course of the whole sequence (see below, Figure 6.16). Using a greater number of switching events should help to reduce the correlation of weakly activated Cy5DNA. The sharp rise in the number of emitting molecule shows that both Cy3/5DNA and Cy*DNA species are strongly photoactivated during each 532 nm pulse.

These experiments show the bulk response of many molecules and therefore provide evidence that psCyDNA can be photoactivated in a reliable manner. The ability to control the switching of psCyDNA is an essential property for its use in optical lock-in detection microscopy.

It is worth noting that recovery from the fluorescent state to the dark state was faster in Cy*DNA, with an average rate constant of 2.7 s^{-1}, compared to 1.5 s^{-1} for Cy3/5DNA. This can perhaps be explained by the fact that incorporating Cy5 in both strands increases the likelihood of interaction with thiols (and thus production of the dark state). Alternatively, photobleaching might be more prevalent in Cy*DNA compared to Cy3/5DNA. A quicker recovery to the dark state means that is it possible to photoswitch at a faster rate. In turn, this means that acquisition times could potentially be reduced; although, this is also dependent on photon yields.

To investigate single-molecule behaviour, intensity traces were recorded for a number of molecules within each sample type. Switching is random in all samples when only 633 nm excitation is used (as shown in Figure 6.14 and Appendix 6, §6.4). Cy5-dCTP molecules show a fairly consistent active (maximum) intensity across all traces. This is consistent with an isolated fluorophore. In contrast, the active intensity of Cy5DNA, Cy3/5DNA, and Cy*DNA molecules is variable. This suggests that an ensemble of Cy5 labels is responsible for the fluorescence signal. In general, the maximum intensity of CyDNA molecules is greater than that of Cy5-dCTP molecules

but, once again, efficient quenching must be occurring to account for the relatively small increase in intensity observed.

Figure 6.16 shows single-molecule intensity traces for Cy5-dCTP, Cy5DNA, Cy3/5DNA, and Cy*DNA molecules under constant illumination at 633 nm and pulses of 532 nm every 5 s (dashed lines). Changes in Cy5DNA molecule intensity show little correlation with 532 nm pulses. In contrast, Cy3/5DNA and Cy*DNA molecules consistently show an increase in intensity with each 532 nm pulse. This indicates that photoswitching behaviour is reliable in single molecules of psCyDNA but not in Cy5DNA.

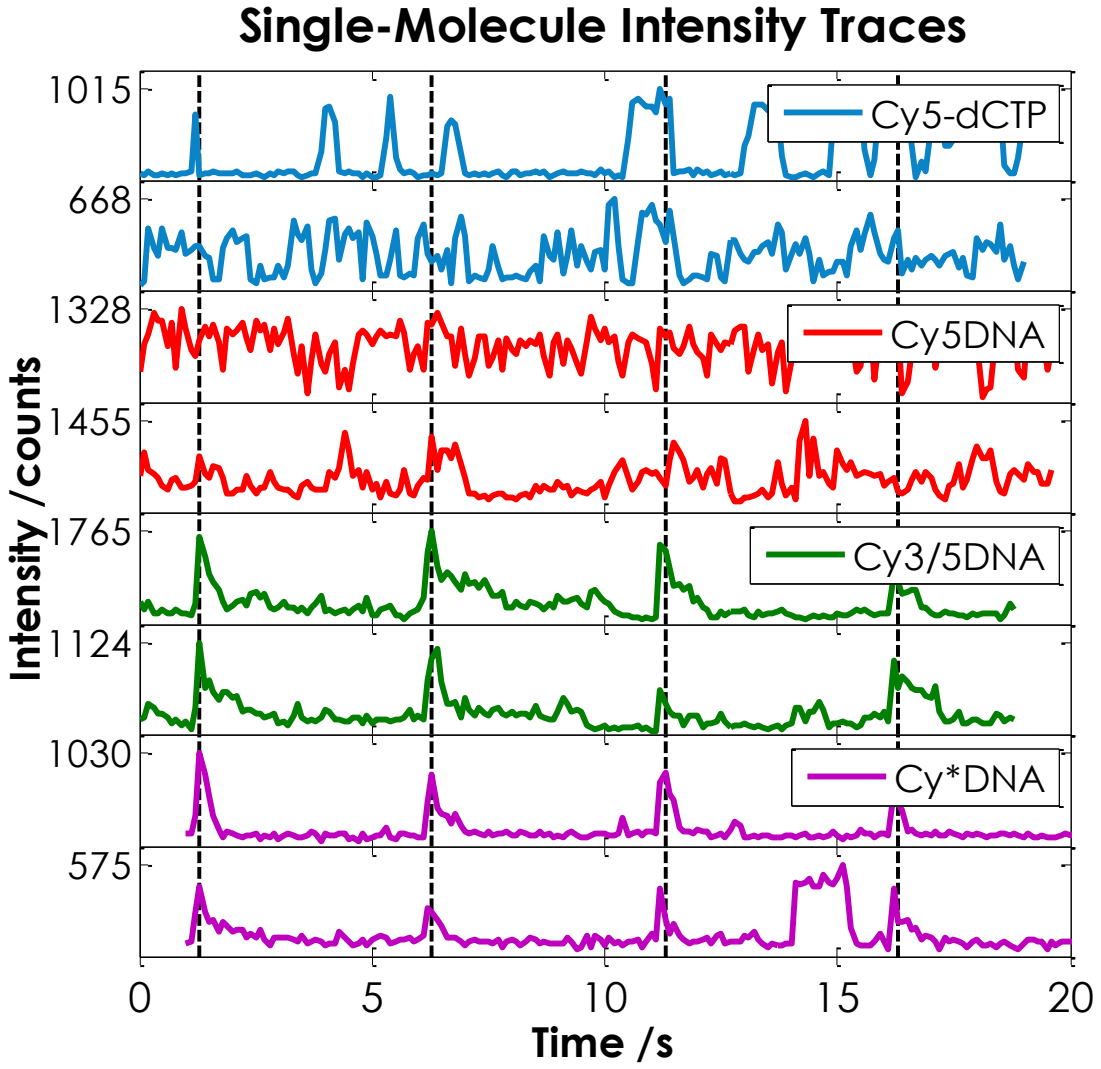

Figure 6.16– Single-molecule intensity traces of Cy5-dCTP, Cy5DNA, Cy3/5DNA, and Cy*DNA molecules under continuous 633 nm excitation and pulses of 532 nm every 5 s (dashed lines). Traces are shown for two of each type of molecule. Traces were created by averaging a 5 × 5 pixel region-of-interest and have been background corrected. The y-axis values represent the maximum intensities for each trace.

It is interesting to note that typical single-molecule intensity traces for Cy3/5DNA and Cy*DNA are quite different. While both molecule types show reliable activation with 532 nm pulses, Cy3/5DNA molecules generally have a higher maximum intensity and show a slower relaxation back to the dark state (consistent with bulk measurements). This suggests that fewer fluorophores are activated in Cy*DNA compared to Cy3/5DNA with each 532 nm pulse. This has implications for super-resolution imaging. As it is necessary to avoid overlapping PSFs, it is highly desirable to have only one fluorophore active within a CyDNA molecule at any given time. Quicker relaxation to the dark state means that the 532 nm pulse repetition rate can be higher, potentially reducing acquisition times. Depending on photon yields, Cy*DNA may allow faster acquisition of super-resolution images than Cy3/5DNA.

In contrast, OLID microscopy requires a large discernible change in intensity at particular moments in time. As such, it is beneficial for as many fluorophores as possible to be activated with each 532 nm pulse. Slower relaxation to the dark state means that scattered light has less correlation to switching behaviour. Therefore, despite requiring longer acquisition times, Cy3/5DNA should be a better sample for OLID imaging than Cy*DNA.

6.3.4 Optical Lock-In Detection Microscopy

Small beads (1.2 μm) were coated with Cy3/5DNA and then mixed with large beads (3.0 μm) coated with a non-switching dye. Figure 6.17 shows a comparison of images produced from this sample. Despite having a much higher mean intensity, the large bead is not visible in the OLID image (Figure 6.17d). In contrast, the smaller photoswitching beads are dramatically enhanced by the OLID analysis. It is also of interest to note that simply taking the standard deviation of individual pixels during the sequence improves the relative contrast of the photoswitching beads compared to the non-switching bead (Figure 6.17c). The resolution in the OLID image has been reduced because the PSFs of the fluorescent molecules are detected across multiple pixels; this increases the correlation coefficient of pixels neighbouring those that actually contain photoswitching molecules.

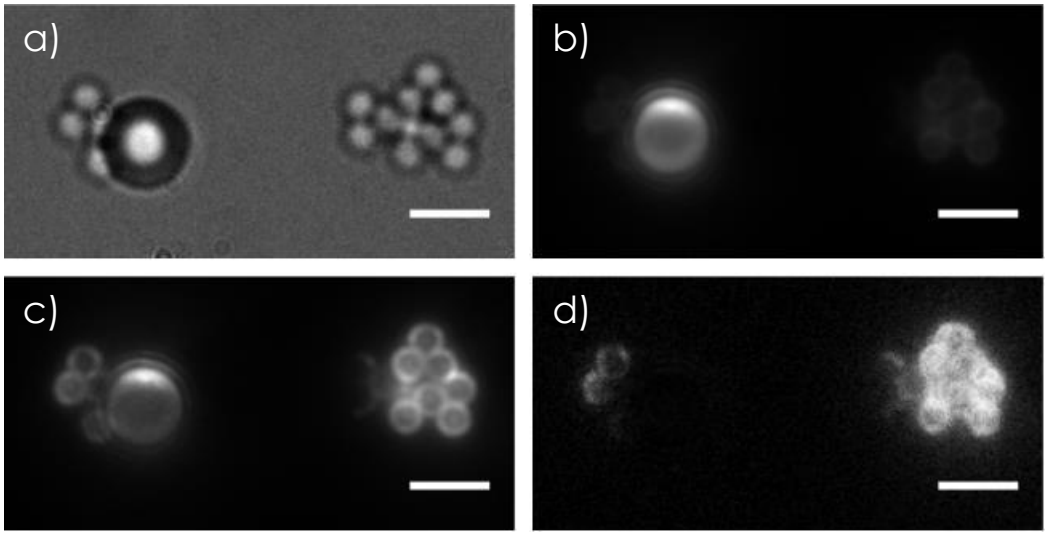

Figure 6.17 – Images of small photoswitching beads and large non-switching beads; a) bright-field, b) mean fluorescence intensity, c) standard deviation, and d) OLID. Scale bars: 3 µm. Enhanced contrast images are shown in Appendix 6, §6.1.3.

Figures 6.18-6.20 show a comparison of mean and OLID images calculated for a sample containing non-switching (N) Cy5DNA and switching (S) Cy3/5DNA.[l] There is a clear improvement in the contrast between switching molecules and background in the OLID image compared to the mean image. A more detailed study was performed on highlighted molecules N1, N2, S1, and S2. Zoomed images and single-pixel traces of these molecules are shown in Figure 6.21 and Figure 6.22, respectively.

[l] While Cy5DNA molecules also switched randomly, switching here refers to photoactivated process enabled by Cy3.

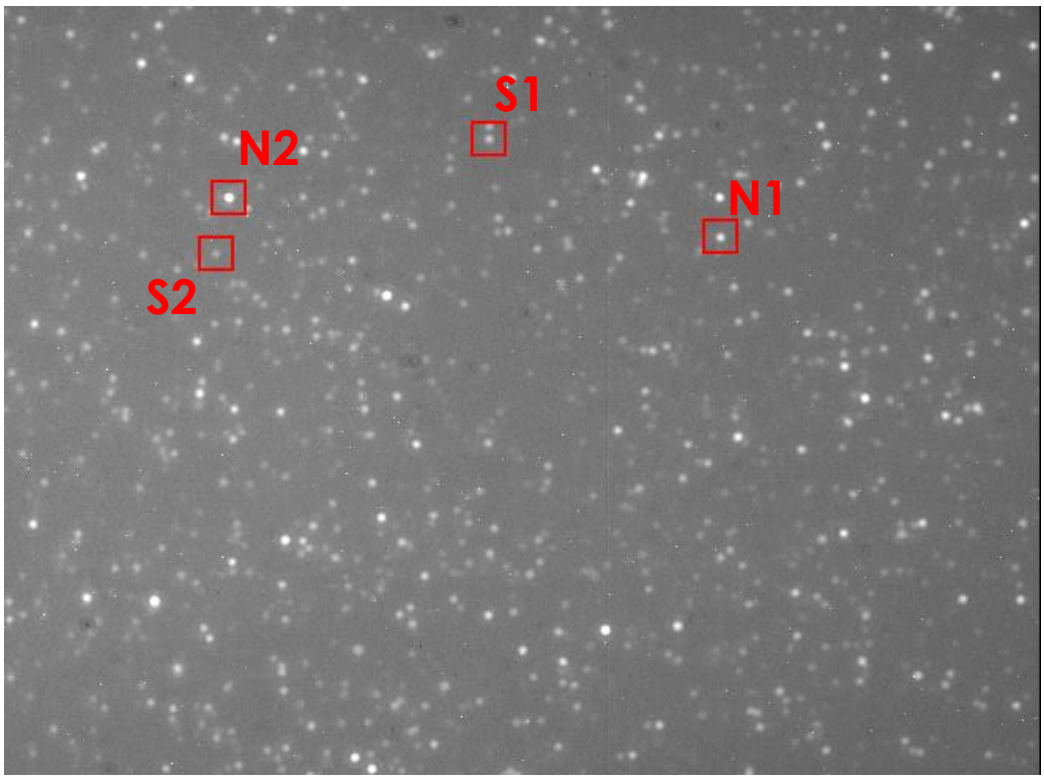

Figure 6.18 – Mean intensity values of a sample containing non-switching (N) Cy5DNA and switching (S) Cy3/5DNA.

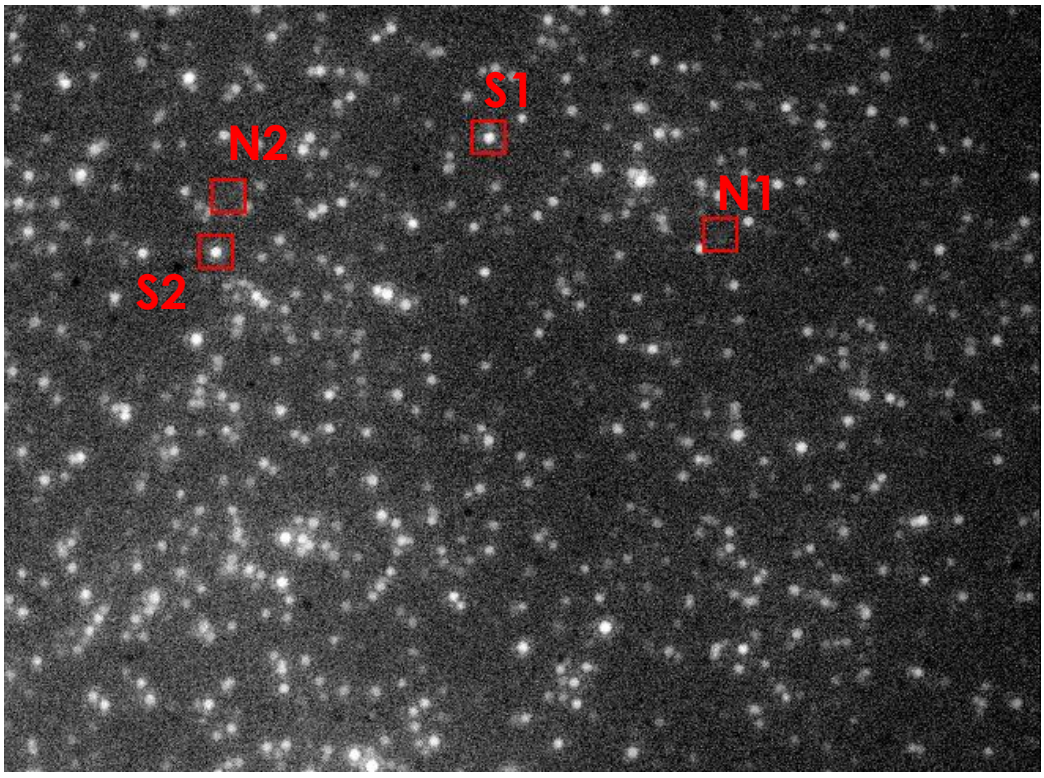

Figure 6.19 –OLID values for the same sample that is shown in Figure 6.18.

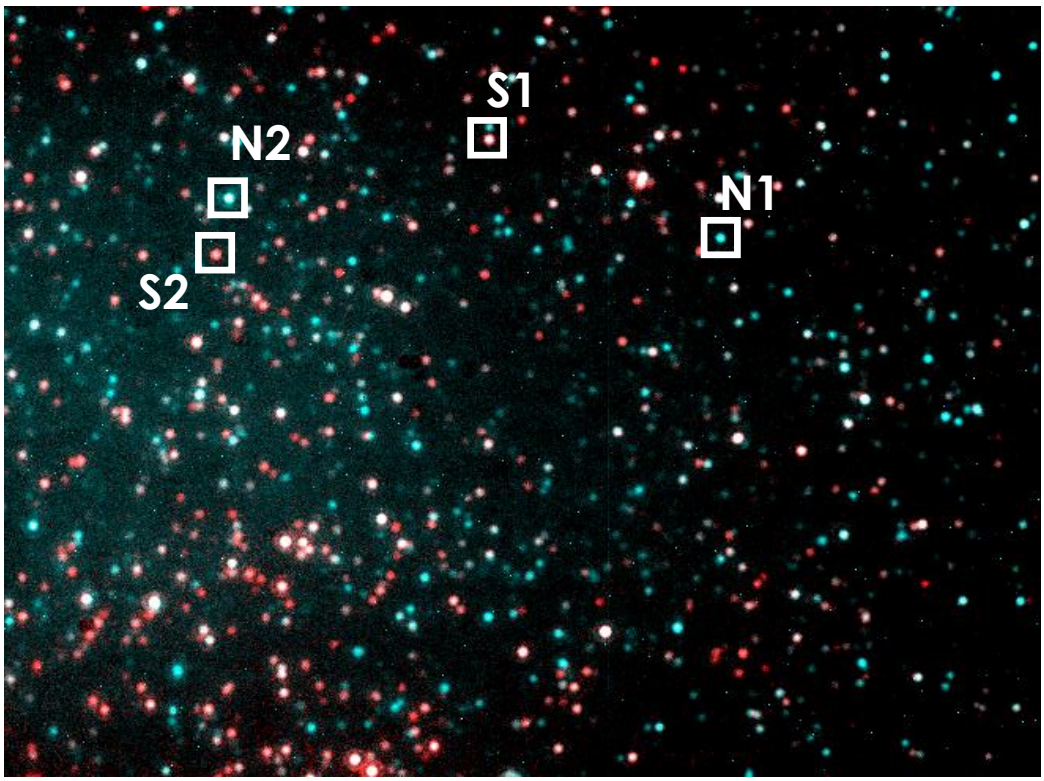

Figure 6.20 – Overlay of mean (cyan, Figure 6.18) and OLID (red, Figure 6.19) images. The image represents a colour coded map where non-switching Cy5DNA molecules are cyan and switching Cy3/5DNA molecules are red. Note that Cy3/5DNA molecules which are visible in the mean image appear white (or pink) in the overlay.

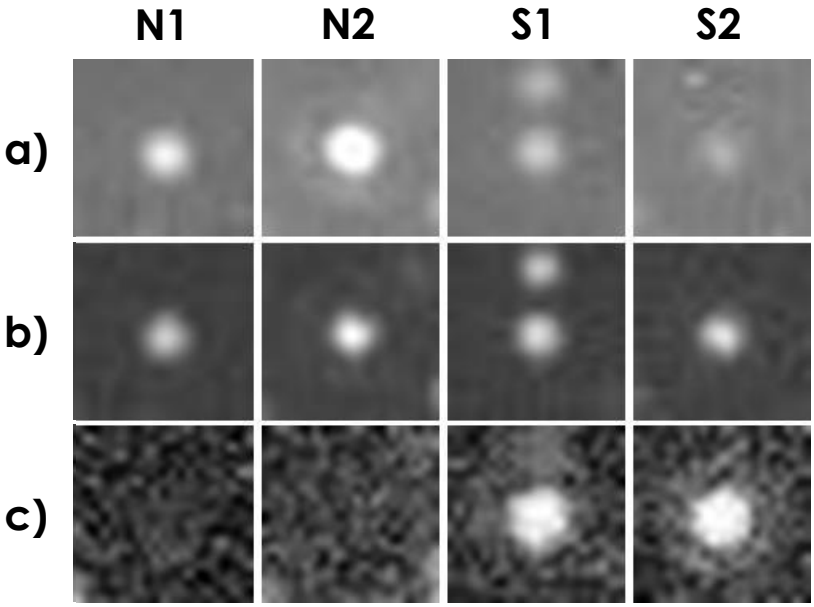

Figure 6.21 – Zoomed images (21 × 21 pixels) of non-switching (N) and switching (S) molecules highlighted in Figures 6.18-6.20: a) mean; b) standard deviation; and c) OLID images. Note that S1 refers to the central molecule in the image.

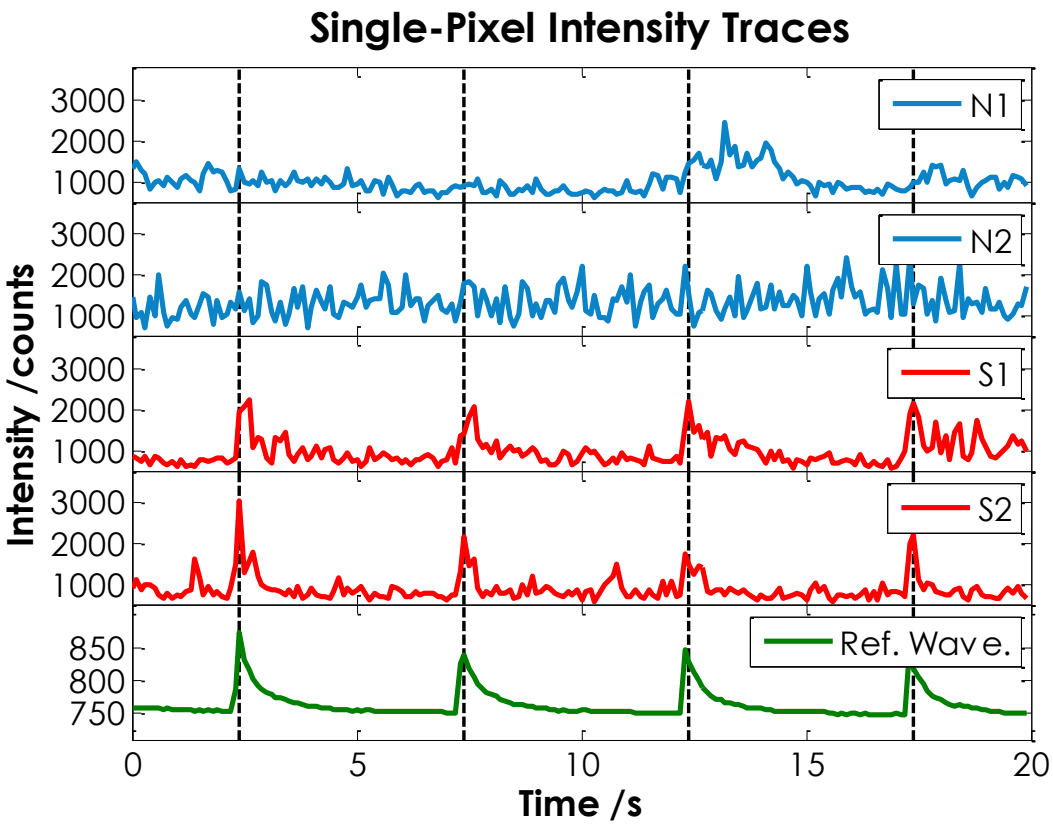

Figure 6.22 – Comparison of single-pixel intensity traces from non-switching (N) Cy5DNA and switching (S) Cy3/5DNA molecules shown in Figure 6.21. The reference waveform (Ref. Wave.) used for OLID analysis is also shown. The dashed lines show the timing of 532 nm pulses.

The reference waveform used for OLID analysis (green trace, Figure 6.22) was created by taking the mean intensity of the entire image over the course of the experiment.[m] Intensity traces for S1 and S2 molecules show correlated switching with 532 nm pulses, identifying them as Cy3/5DNA. N1 and N2 intensity traces have randomly fluctuating intensity and so were identified as Cy5DNA molecules. The consequence of this contrasting behaviour is evident in the expanded images shown in Figure 6.21. In the mean image (Figure 6.21a) Cy5DNA molecules (N1 and N2) are bright and highly visible whereas they are undetected (absent) in the OLID correlation image (Figure 6.21c). The situation is almost the complete opposite for Cy3/5DNA

[m] As many molecules in the sample were photoactivated at the same time, it was also possible to use the number of emitting molecules as a reference waveform. Correlation between the number of emitters and the mean intensity was very high (0.981). Details of this comparison can be found in Appendix 6, §6.1.2.

molecules, which have increased intensity in the OLID image. Thus, it has been possible to discriminate between high levels of noise (Cy5DNA in this sample) and the signal from the probe (Cy3/5DNA) using OLID analysis. Table 6.4 summarises the mean intensity and correlation coefficients for the central pixel in the expanded images of Figure 6.21. Values confirm the enhancement to the relative intensity of switching molecules compared to non-switching molecules. While mean intensity values are greater for non-switching (N) molecules, only switching (S) molecules have a high correlation coefficient.

Molecule	**N1**	**N2**	**S1**	**S2**
Mean Intensity /counts per 100 ms	995	1327	952	878
OLID Correlation Coefficient	0.197	0.148	0.758	0.782

Table 6.4 – Comparison of mean intensities and correlation values for the molecules highlighted in Figures 6.18-6.20. Values were calculated from the central pixel of the images shown in Figure 6.21.

It is again interesting to note that improvement can be made to image contrast by using standard deviation values, as observed in Figure 6.21b. This is somewhat expected given that the intensity of a switching molecule will likely fluctuate more than noise. While it is not possible to distinguish between Cy5DNA and Cy3/5DNA molecules in this case, using standard deviation values could potentially improve image contrast for more conventional fluorophores.

The results above show that psCyDNA is an excellent sample to use for OLID microscopy. Given the aforementioned advantages of psCyDNA, the combination of psCyDNA and OLID could be used in a variety of future studies to great success.

It is important to note that in this OLID experiment only four switching events were used. This is far fewer than the number of events used in the study by Marriott *et al.*[3] Presumably, the large number of switches was required because a fairly long scan time (1 s) was used for each confocal image. This meant few data points were collected in each photoswitching cycle. In contrast, using wide-field detection in the present study allowed images to be recorded 10 times faster, primarily because many points could be collected in each cycle. Wide-field OLID microscopy could therefore be used to obtain high contrast images in short periods of time. Shorter acquisition times mean that artefacts from changes in the sample due to, for example, sample drift are reduced.

A brief discussion about the detrimental effects of photobleaching to the OLID analysis method can be found in Appendix 6, §6.1.4.

6.3.5 Super-Resolution Imaging of Cy3/5DNA

Figure 6.23 shows the raw localisations and the reconstructed super-resolution image of a single Cy3/5DNA molecule. It should be noted that the experimental data presented was supplied by Dr. Cristina Flors.[125]

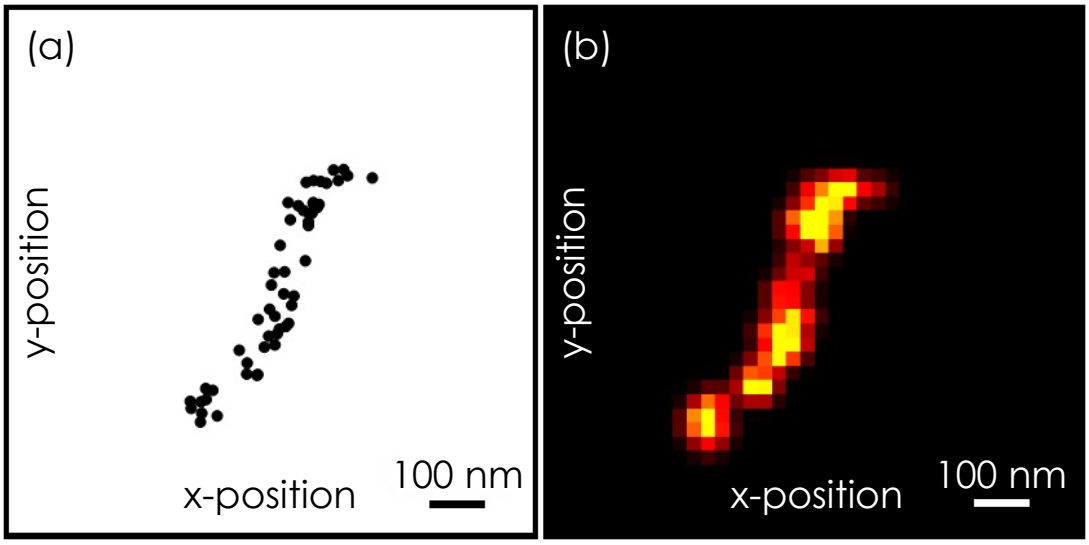

Figure 6.23 – Super-resolution (PALM/STORM) image a single Cy3/5DNA molecule spin coated onto a coverslip: (a) raw localisations and (b) reconstruction. The visible length of the DNA is fairly consistent with the expected number of base pairs (1.3 kbp, ~440 nm).

In addition to showing the general structure of the substrate, localisations show that dyes have been incorporated throughout the whole DNA duplex. The distribution of dyes does not appear to be completely homogeneous; however, there is the possibility that energy transfer to dyes in a lower energy state might cause apparent dark areas. The relatively large width of the DNA (~70 nm) could be due to the DNA not being firmly attached to the coverslip and also the linker between deoxycytidine and the Cy5 chromophore. Localisation error was on the order of 20-30 nm.

The final section of this chapter will attempt to bring together an overall picture of the results of this investigation and aim to provide an outlook into potential future studies and developments related to super-resolution imaging of DNA.

6.4 Conclusions

CyDNA, a densely labelled, highly fluorescent biopolymer, has been investigated to determine its suitability for use in fluorescence microscopy applications, particularly where reliable photoswitching is required.

Photophysical studies have shown that dye-dye interactions, such as the formation of H-aggregates, compromise the brightness of CyDNA molecules. This could hinder the use of CyDNA in future applications, such as fluorescence *in situ* hybridisation (FISH)[16] and in microarray and microfluidic devices.[126] Investigation into labelling patterns may provide an optimal dye density to alleviate quenching effects. Randolph *et al.*[121] have already performed a study to this effect for the Cy5 dye; however, the substrates used were of very short length compared to CyDNA. Also, only one strand in the duplex was labelled, which meant that interstrand dye interactions could not be taken into account. More rigorous studies are therefore required.

Cy3/5DNA and Cy*DNA (psCyDNA) have been shown to be highly efficient photoswitching macromolecules. This holds great promise for the future use of this material as photoswitching is essential for localisation-based super-resolution studies. The high density of labels also means that psCyDNA can be used to image in fine detail. Thus, even nanoscale structures could be visualised in very high resolution using this biopolymer.

Future super-resolution studies may help to elucidate other quenching mechanisms present in CyDNA. Given the similarity of the two systems, it seems plausible that low-energy traps evident in conjugated polymers would also exist in CyDNA. If low-energy traps do exist in CyDNA then it may be possible to see them 'jumping' around the duplex structure with super-resolution techniques. This 'jumping' could happen if the traps are sequentially photobleached or if there is some conformational change, such as folding, in the DNA during acquisition.

Labelling in CyDNA is sequence-specific as it only occurs at natural cytidine sites. CyDNA therefore potentially has a great advantage over intercalating dyes for studies involving DNA because it is, in principle at least, possible to create a probe which contains dyes at specific locations (and thus specific separations) within the substrate. This could be very useful for determining the dimensions of a structure as the distance between dyes would be known before imaging was performed. In combination with DNA origami, this could provide an elegant method for the creation of reference standards for microscope image dimensions.[127] This would, however, require an

appropriate PCR template to be developed and, in addition, the cyanine dyes may also disrupt the compact origami structure; this may diminish the precision of the distances between labels.

The reliance on PCR to construct CyDNA means that the true labelling density can only be estimated from the percentage of labelled bases used in the reaction mixture. Super-resolution studies on stretched CyDNA might allow the real density of dyes to be estimated and also how homogeneous the labelling is along the substrate. Given the above discussion about dye-dye interactions and optimisation of label density, this information would be very useful to obtain.

OLID correlation analysis has been shown to discriminate successfully between Cy5DNA and Cy3/5DNA, despite the former generally having higher intensity. This bodes well for the use of OLID analysis in future studies where unwanted signal (or background) is at the same order of magnitude as the probe signal. This means that only very small concentrations of a probe molecule would be necessary to image a sample. This could be particularly beneficial for biological samples where it may be difficult to introduce probes or where the probe might have toxicological effects at high concentrations.

Super-resolution optical fluctuation imaging (SOFI) is a recently developed method of data analysis that allows the improvement of both contrast and resolution.[128–130] The results from the present work show that simply taking the standard deviation of each pixel value can improve contrast between fluorophore emission and background signal (see Figure 6.17 and Figure 6.21). This is by virtue of the fact that fluorescence intensity will fluctuate more than the background signal. SOFI appears to be a much more rigorous extension of this idea and has the added advantage that it also improves resolution. It would be interesting to see if the controlled optical modulation used in OLID could be used in combination with the SOFI analysis method to further improve image quality.

An obvious possibility for the future would be to replace the other natural bases (A, G, and T) with labelled analogues. Ideally, if one could incorporate a different fluorophore into each base then it would be possible to determine the conformational geometry of the DNA whilst also obtaining its base sequence. While easy to envision, this would likely be very difficult to achieve; although, it does appear to be feasable.[94]

It would be interesting to see if it was possible to use the dSTORM[62] method to image Cy5DNA. If this were possible, then it would negate the need for Cy3 fluorophores. During experiments only Cy5 molecules were imaged and so replacement

of Cy3 dyes with Cy5 would increase the density of detectable labels, improving the potential resolution of the images produced; however, taking away Cy3 from the system would remove the controllable nature of the photoswitching in the sample, meaning that OLID analysis would not be possible. It is also likely that the severe fluorescence quenching currently observed in Cy5DNA would prohibit dSTORM imaging and so it may be necessary to optimise labelling strategies before applying this method.

Competition between the various super-resolution microscopy techniques which exist can only be seen as a good thing and has likely been one of the main contributing factors driving the rapid advancement of this field over the past couple of decades; even sub-molecule-scale studies have recently been predicted.[131]

Despite this progress and optimism, the lack of suitable labelling strategies, particularly for DNA, is a significant obstacle in the path of realising the full potential of super-resolution microscopy and other advanced imaging techniques, such as OLID microscopy. Photophysical studies of novel systems, such as the one presented in this chapter, are therefore crucial to develop microscopy further. Surpassing current limitations will allow important questions to be answered and, in turn, improve our knowledge of the nanoscale world, a place of great interest to many subject areas.

6.5 References

1. C. Flors and W. C. Earnshaw, *Curr. Opin. Chem. Biol.*, 2011, **15**, 838–44.

2. A. Fürstenberg and M. Heilemann, *Phys. Chem. Chem. Phys.*, 2013, **15**, 14919–30.

3. G. Marriott, S. Mao, T. Sakata, J. Ran, D. K. Jackson, C. Petchprayoon, T. J. Gomez, E. Warp, O. Tulyathan, H. L. Aaron, E. Y. Isacoff, and Y. Yan, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 17789–94.

4. Y. Yan, M. E. Marriott, C. Petchprayoon, and G. Marriott, *Biochem. J.*, 2011, **433**, 411–22.

5. Y. Yan, C. Petchprayoon, S. Mao, and G. Marriott, *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 2013, **368**, 20120031.

6. Q. Wei and A. Wei, *Chemistry*, 2011, **17**, 1080–91.

7. B. Huang, H. Babcock, and X. Zhuang, *Cell*, 2010, **143**, 1047–58.

8. C. Flors, C. N. J. Ravarani, and D. T. F. Dryden, *Chemphyschem*, 2009, **10**, 2201–4.

9. C. Flors, *Photochem. Photobiol. Sci.*, 2010, **9**, 643–8.

10. I. Schoen, J. Ries, E. Klotzsch, H. Ewers, and V. Vogel, *Nano Lett.*, 2011, **11**, 4008–11.

11. R. Jungmann, C. Steinhauer, M. Scheible, A. Kuzyk, P. Tinnefeld, and F. C. Simmel, *Nano Lett.*, 2010, **10**, 4756–61.

12. A. Benke and S. Manley, *Chembiochem*, 2012, **13**, 298–301.

13. P. J. M. Zessin, K. Finan, and M. Heilemann, *J. Struct. Biol.*, 2012, **177**, 344–8.

14. C. Vranken, J. Deen, L. Dirix, T. Stakenborg, W. Dehaen, V. Leen, J. Hofkens, and R. K. Neely, *Nucleic Acids Res.*, 2014, 1–10.

15. R. Jungmann, M. S. Avendaño, J. B. Woehrstein, M. Dai, W. M. Shih, and P. Yin, *Nat. Methods*, 2014, **11**, 313–8.

16. N. Ramsay, A.-S. Jemth, A. Brown, N. Crampton, P. Dear, and P. Holliger, *J. Am. Chem. Soc.*, 2010, **132**, 5096–104.

17. M. J. Rust, M. Bates, and X. Zhuang, *Nat. Methods*, 2006, **3**, 793–5.

18. N. R. Conley, J. S. Biteen, and W. E. Moerner, *J. Phys. Chem. B*, 2008, **112**, 11878–80.

19. M. Bates, T. Blosser, and X. Zhuang, *Phys. Rev. Lett.*, 2005, **94**, 108101.

20. W. E. Moerner and L. Kador, *Phys. Rev. Lett.*, 1989, **62**, 2535–8.

21. W. P. Ambrose and W. E. Moerner, *Nature*, 1991, **349**, 225–7.

22. M. Orrit and J. Bernard, *Phys. Rev. Lett.*, 1990, **65**, 2716–9.

23. L. Bai, T. J. Santangelo, and M. D. Wang, *Annu. Rev. Biophys. Biomol. Struct.*, 2006, **35**, 343–60.

24. F. Zernike, *Physica*, 1942, **IX**, 686–98.

25. F. Zernike, *Physica*, 1942, **IX**, 974–86.

26. F. Zernike, *Science*, 1955, **121**, 345–9.

27. H. E. Rosenberger, in *Interpretive Techniques for Microstructural Analysis*, eds. J. L. McCall and P. M. French, Plenum Press, New York, 1977, pp. 79–104.

28. D. J. Stephens and V. J. Allan, *Science*, 2003, **300**, 82–6.

29. M. Minsky, *US Pat. 3,013,467*, 1961.

30. O. Shimomura, F. H. Johnson, and Y. Saiga, *J. Cell. Comp. Physiol.*, 1962, **59**, 223–39.

31. R. B. Mujumdar, L. A. Ernst, S. R. Mujumdar, C. J. Lewis, and A. S. Waggoner, *Bioconjug. Chem.*, 1993, **4**, 105–11.

32. S. R. Mujumdar, R. B. Mujumdar, C. M. Grant, and A. S. Waggoner, *Bioconjug. Chem.*, 1996, **7**, 356–62.

33. M. Bates, B. Huang, and X. Zhuang, *Curr. Opin. Chem. Biol.*, 2008, **12**, 505–14.

34. G. B. Airy, in *Transactions of the Cambridge Philosophical Society*, Cambridge University Press, 1835, pp. 283–91.

35. Lord Rayleigh, *Philos. Mag.*, 1879, **8**, 261–74.

36. S. T. Hess, T. P. K. Girirajan, and M. D. Mason, *Biophys. J.*, 2006, **91**, 4258–72.

37. R. Schmidt, C. A. Wurm, A. Punge, A. Egner, S. Jakobs, and S. W. Hell, *Nano Lett.*, 2009, **9**, 2508–10.

38. G. Binnig, C. F. Quate, and C. Gerber, *Phys. Rev. Lett.*, 1986, **56**, 930–4.

39. L. Gross, F. Mohn, N. Moll, B. Schuler, A. Criado, E. Guitian, D. Pena, A. Gourdon, and G. Meyer, *Science*, 2012, **337**, 1326–9.

40. D. G. de Oteyza, P. Gorman, Y.-C. Chen, S. Wickenburg, A. Riss, D. J. Mowbray, G. Etkin, Z. Pedramrazi, H.-Z. Tsai, A. Rubio, M. F. Crommie, and F. R. Fischer, *Science*, 2013, **29**, 1219–27.

41. J. Zhang, P. Chen, B. Yuan, W. Ji, Z. Cheng, and X. Qiu, *Science*, 2013, **100**, 1–7.

42. G. Binnig and H. Rohrer, *US Pat. 4,343,993*, 1982, 1–11.

43. G. Binnig and H. Rohrer, *IBM J. Res. Dev.*, 1986, **30**, 279–93.

44. G. J. Simpson, S. W. L. Hogan, M. Caffio, C. J. Adams, H. Früchtl, T. van Mourik, and R. Schaub, *Nano Lett.*, 2014, **14**, 634–9.

45. M. von Ardenne, *UK Pat. GB 511,204-A*, 1938.

46. M. von Ardenne, *Zeitschrift für Phys.*, 1938, **108**, 553–72.

47. J. B. Pendry, *Phys. Rev. Lett.*, 2000, **85**, 3966–9.

48. Z. Wang, W. Guo, L. Li, B. Luk'yanchuk, A. Khan, Z. Liu, Z. Chen, and M. Hong, *Nat. Commun.*, 2011, **2**, 218.

49. E. A. Synge, *Philos. Mag.*, 1928, **6**, 356–62.

50. E. A. Ash and G. Nicholls, *Nature*, 1972, **237**, 510–2.

51. D. W. Pohl, W. Denk, and M. Lanz, *Appl. Phys. Lett.*, 1984, **44**, 70651–3.

52. M. G. L. Gustafsson, *J. Microsc.*, 2000, **198**, 82–7.

53. M. G. L. Gustafsson, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 13081–6.

54. E. H. Rego, L. Shao, J. J. Macklin, L. Winoto, G. A. Johansson, N. Kamps-Hughes, M. W. Davidson, and M. G. L. Gustafsson, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, E135–43.

55. K. Wicker and R. Heintzmann, *J. Opt.*, 2010, **12**, 084010.

56. S. W. Hell and J. Wichmann, *Opt. Lett.*, 1994, **19**, 780–2.

57. T. A. Klar and S. W. Hell, *Opt. Lett.*, 1999, **24**, 954–6.

58. A. Chmyrov, J. Keller, T. Grotjohann, M. Ratz, E. D'Este, S. Jakobs, C. Eggeling, and S. W. Hell, *Nat. Methods*, 2013, **10**, 737–40.

59. E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, *Science*, 2006, **313**, 1642–5.

60. R. Henriques, M. Lelek, E. F. Fornasiero, F. Valtorta, C. Zimmer, and M. M. Mhlanga, *Nat. Methods*, 2010, **7**, 339–40.

61. A. G. York, A. Ghitani, A. Vaziri, M. W. Davidson, and H. Shroff, *Nat. Methods*, 2011, **8**, 327–33.

62. M. Heilemann, S. van de Linde, M. Schüttpelz, R. Kasper, B. Seefeldt, A. Mukherjee, P. Tinnefeld, and M. Sauer, *Angew. Chemie, Int. Ed.*, 2008, **47**, 6172–6.

63. S. van de Linde, A. Löschberger, T. Klein, M. Heidbreder, S. Wolter, M. Heilemann, and M. Sauer, *Nat. Protoc.*, 2011, **6**, 991–1009.

64. G. Shtengel, J. A. Galbraith, C. G. Galbraith, J. Lippincott-Schwartz, J. M. Gillette, S. Manley, R. Sougrat, C. M. Waterman, P. Kanchanawong, M. W. Davidson, R. D. Fetter, and H. F. Hess, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 3125–30.

65. B. Huang, W. Wang, M. Bates, and X. Zhuang, *Science*, 2008, **319**, 810–3.

66. D. Baddeley, M. B. Cannell, and C. Soeller, *Microsc. Microanal.*, 2010, **16**, 64–72.

67. A. Mustafa, *Chem. Rev.*, 1948, **43**, 509–23.

68. E. Fisher and Y. Hirshberg, *J. Chem. Soc.*, 1952, 4522–4.

69. R. Nakao, F. Noda, T. Horii, and Y. Abe, *Polym. Adv. Technol.*, 2002, **13**, 81–6.

70. R. Ando, H. Hama, M. Yamamoto-Hino, H. Mizuno, and A. Miyawaki, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 12651–6.

71. D. M. Chudakov, V. V. Belousov, A. G. Zaraisky, V. V. Novoselov, D. B. Staroverov, D. B. Zorov, S. Lukyanov, and K. A. Lukyanov, *Nat. Biotechnol.*, 2003, **21**, 191–4.

72. G. H. Patterson and J. Lippincott-Schwartz, *Science*, 2002, **297**, 1873–7.

73. U. Endesfelder, S. Malkusch, B. Flottmann, J. Mondry, P. Liguzinski, P. J. Verveer, and M. Heilemann, *Molecules*, 2011, **16**, 3106–18.

74. K. A. Lukyanov, D. M. Chudakov, S. Lukyanov, and V. V. Verkhusha, *Nat. Rev. Mol. Cell Biol.*, 2005, **6**, 885–91.

75. R. Ando, H. Mizuno, and A. Miyawaki, *Science*, 2004, **306**, 1370–3.

76. M. Heilemann, S. van de Linde, A. Mukherjee, and M. Sauer, *Angew. Chemie, Int. Ed.*, 2009, **48**, 6903–8.

77. J. Martini, K. Schmied, R. Palmisano, K. Toensing, D. Anselmetti, and T. Merkle, *J. Struct. Biol.*, 2007, **158**, 401–9.

78. G. Berkovic, V. Krongauz, and V. Weiss, *Chem. Rev.*, 2000, **100**, 1741–54.

79. X. Li, J. Li, Y. Wang, T. Matsuura, and J. Meng, *J. Photochem. Photobiol. A Chem.*, 2004, **161**, 201–13.

80. E. Fron, C. Flors, G. Schweitzer, S. Habuchi, H. Mizuno, R. Ando, F. C. De Schryver, A. Miyawaki, and J. Hofkens, *J. Am. Chem. Soc.*, 2007, **129**, 4870–1.

81. M. M. Warren, M. Kaucikas, A. Fitzpatrick, P. Champion, J. T. Sage, and J. J. van Thor, *Nat. Commun.*, 2013, **4**, 1461.

82. C. Duan, V. Adam, M. Byrdin, and D. Bourgeois, in *Photoswitching Proteins: Methods and Protocols*, ed. S. Cambridge, Springer, New York, NY, 2014, vol. 1148, pp. 177–202.

83. S. van de Linde, I. Krstić, T. Prisner, S. Doose, M. Heilemann, and M. Sauer, *Photochem. Photobiol. Sci.*, 2011, **10**, 499–506.

84. J. Vogelsang, T. Cordes, C. Forthmann, C. Steinhauer, and P. Tinnefeld, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 8107–12.

85. R. E. Thompson, D. R. Larson, and W. W. Webb, *Biophys. J.*, 2002, **82**, 75–83.

86. G. Patterson, M. Davidson, S. Manley, and J. Lippincott-Schwartz, *Annu. Rev. Phys. Chem.*, 2010, **61**, 345–67.

87. C. E. Shannon, *Proc. IRE*, 1949, 10–21.

88. H. Sies and C. F. M. Menck, *Mutat. Res.*, 1992, **275**, 367–75.

89. C. Flors, *Biopolymers*, 2011, **95**, 290–7.

90. A. Salic and T. J. Mitchison, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 2415–20.

91. S. Mao, R. K. P. Benninger, Y. Yan, C. Petchprayoon, D. Jackson, C. J. Easley, D. W. Piston, and G. Marriott, *Biophys. J.*, 2008, **94**, 4515–24.

92. G. Giller, T. Tasara, B. Angerer, K. Mühlegger, M. Amacker, and H. Winter, *Nucleic Acids Res.*, 2003, **31**, 2630–5.

93. T. Tasara, B. Angerer, M. Damond, H. Winter, S. Dörhöfer, U. Hübscher, and M. Amacker, *Nucleic Acids Res.*, 2003, **31**, 2636–46.

94. J. P. Anderson, B. Angerer, and L. A. Loeb, *Biotechniques*, 2005, **38**, 257–64.

95. G. Stengel, M. Urban, B. W. Purse, and R. D. Kuchta, *Anal. Chem.*, 2009, **81**, 9079–85.

96. G. T. Dempsey, M. Bates, W. E. Kowtoniuk, D. R. Liu, R. Y. Tsien, and X. Zhuang, *J. Am. Chem. Soc.*, 2009, **131**, 18192–3.

97. M. Heilemann, E. Margeat, R. Kasper, M. Sauer, and P. Tinnefeld, *J. Am. Chem. Soc.*, 2005, **127**, 3801–6.

98. M. Kasha, *J. Chem. Phys.*, 1952, **20**, 71–4.

99. J. Widengren and P. Schwille, *J. Phys. Chem. A*, 2000, **104**, 6416–28.

100. Z. Huang, D. Ji, S. Wang, A. Xia, F. Koberling, M. Patting, and R. Erdmann, *J. Phys. Chem. A*, 2006, **110**, 45–50.

101. W. R. Ware, *J. Phys. Chem.*, 1962, 1949–52.

102. F. Wilkinson, *Pure Appl. Chem.*, 1997, **69**, 851–6.

103. M. K. Singh, *Phys. Chem. Chem. Phys.*, 2009, **11**, 7225–30.

104. P. Tinnefeld, D.-P. Herten, and M. Sauer, *J. Phys. Chem. A*, 2001, **105**, 7989–8003.

105. M. E. Sanborn, B. K. Connolly, K. Gurunathan, and M. Levitus, *J. Phys. Chem. B*, 2007, **111**, 11064–74.

106. A. Iqbal, S. Arslan, B. Okumus, T. J. Wilson, G. Giraud, D. G. Norman, T. Ha, and D. M. J. Lilley, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 11176–81.

107. P. Holliger, personal communication, MRC Laboratory of Molecular Biology, Cambridge, United Kingdom, 2011.

108. G. Stengel, M. Urban, B. W. Purse, and R. D. Kuchta, *Anal. Chem.*, 2010, **82**, 1082–9.

109. H. J. Gruber, C. D. Hahn, G. Kada, C. K. Riener, G. S. Harms, W. Ahrer, T. G. Dax, and H.-G. Knaus, *Bioconjug. Chem.*, 2000, **11**, 696–704.

110. Andor Solis, *Version 4.8.3*, Andor Technology Ltd., Springvale Business Park, Belfast, United Kingdom, See; http://www.andor.com/, 2012.

111. LabVIEW, *Version 2012*, National Instruments Corporation, Austin, Texas, United States of America, See; http://www.ni.com/labview/, 2012.

112. IGOR Pro, *Version 6.2.2*, WaveMetrics, Inc., Portland, Oregon, United States of America, See; http://www.wavemetrics.com/, 2012.

113. P. Dedecker, personal communication, Katholieke Universiteit, Leuven, Belgium, See; http://www.igorexchange.com/project/Localizer, 2011.

114. Microsoft Excel (2010), *Version 14.0*, Microsoft Corporation, Microsoft Redmond Campus, Redmond, Washington, United States of America, 2010.

115. Java, *Version JDK 6u23*, Oracle Corporation, Redwood Shores, Redwood City, California, United States of America, See; http://www.oracle.com/technetwork/java/, 2012.

116. W. S. Rasband, *ImageJ Version 1.44p*, U. S. National Institutes of Health, Bethesda, Maryland, United States of America, See; http://imagej.nih.gov/ij/, 2012.

117. M. D. Abràmoff, P. J. Magalhães, and S. J. Ram, *Biophotonics Int.*, 2004, **11**, 36–42.

118. C. A. Schneider, W. S. Rasband, and K. W. Eliceiri, *Nat. Methods*, 2012, **9**, 671–5.

119. MATLAB, *Version 8.2.0 (R2013b)*, The MathWorks, Inc., Natick, Massachusetts, United States of America, 2013.

120. J. Kang and O. Kaczmarek, *Int. J. Polym. Sci.*, 2010, **2010**, 264781.

121. J. B. Randolph and A. S. Waggoner, *Nucleic Acids Res.*, 1997, **25**, 2923–9.

122. K. C. Hannah and B. A. Armitage, *Acc. Chem. Res.*, 2004, **37**, 845–53.

123. M. Wang, G. L. Silva, and B. A. Armitage, *J. Am. Chem. Soc.*, 2000, **41**, 9977–86.

124. P. F. Barbara, A. J. Gesquiere, S.-J. Park, and Y. J. Lee, *Acc. Chem. Res.*, 2005, **109**, 602–10.

125. C. Flors, personal communication, University of Edinburgh, Edinburgh, United Kingdom, 2012.

126. P. A. C. 't Hoen, F. de Kort, G. J. B. van Ommen, and J. T. den Dunnen, *Nucleic Acids Res.*, 2003, **31**, e20 1–8.

127. C. Steinhauer, R. Jungmann, T. L. Sobey, F. C. Simmel, and P. Tinnefeld, *Angew. Chemie, Int. Ed.*, 2009, **48**, 8870–3.

128. T. Dertinger, R. Colyer, G. Iyer, S. Weiss, and J. Enderlein, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 22287–92.

129. T. Dertinger, R. Colyer, R. Vogel, J. Enderlein, and S. Weiss, *Opt. Express*, 2010, **18**, 18875–85.

130. P. Dedecker, G. C. H. Mo, T. Dertinger, and J. Zhang, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 10909–14.

131. S. W. Hell, *Nat. Methods*, 2008, **6**, 24–32.

Chapter 7: Conclusions

An expert is a person who has found out by his own painful experience all the mistakes that one can make in a very narrow field.

– Niels Bohr

The main aim of this thesis has been to provide new insights into the structure and dynamics of DNA. This has been achieved through the use of fluorescence-based techniques and by performing computational calculations. This chapter will summarise the most significant findings and also provide a brief outlook to possible future experiments and developments.

Time-resolved fluorescence experiments have been performed to investigate relaxation dynamics in 2AP-containing dinucleotides in various alcohol-water mixtures. Initial measurements on the free 2AP fluorophore showed that solvent polarity has significant influence on the observed fluorescence lifetime; this is consistent with the currently favoured interpretation that suggests solvation affects the barrier height to intersystem crossing, which is a competitive pathway to fluorescence emission. The use of the dinucleotide construct offered a simplified model of DNA and allowed base-specific effects to be isolated. The most rapid decay component was found to be shortest and the most dominant for 2AP-G while being longest and least dominant for 2AP-A; the overall trend for the fractional contribution of the short component was G > T > C > I > A (the order was reversed for the magnitude of the associated lifetime). The observed trends could be explained by consideration of charge transfer efficiency in these dinucleotides. With the aim of perturbing the distribution of conformational states present, the solvent environment of the 2AP-N dinucleotides was altered by adding alcohol. The findings further corroborate the importance of structural dynamics in determining the complex fluorescence behaviour observed in these systems. The overall behaviour of the decay A-factors suggested that there was an interrelationship between A_1, A_3, and A_4 but not A_2. The independence of A_2 might indicate that this component corresponds to a distinct relaxation mechanism from the others. A particularly interesting observation was the dramatic switch in populations of A_3 and A_4 components after an alcohol content threshold had been reached. The implication of this behaviour is that there is a solvent composition (most likely defined by a specific polarity) beyond which the open conformation of the dinucleotide becomes significantly more stable (or is accessed more readily). It would be fascinating to see if there were some base-specific influence on the threshold; this may provide some insight into the relative stacking energy of the different bases. Unfortunately, the compositions used in this study were too coarsely separated to gain any detailed insight into this issue and all dinucleotides changed their conformational distribution between 30% and 50% (w/w) glycerol content. An obvious extension of this work

would be to investigate a more diverse range of solvents; especially those with different H-bonding ability.

Analysis of the fluorescence decays measured for the 2AP-N dinucleotides was quite challenging, particularly for the 2AP-A dinucleotide, which appeared to exhibit unusual behaviour. This highlighted a more general problem in the analysis of time-resolved fluorescence measurements; specifically, that there often can be uncertainty about the accuracy of the model that is used to fit the decay, even when the model provides an adequate description of the data. To try and address this issue, various time-resolved fluorescence analysis methods have been investigated. Real and simulated decays were used to evaluate the relative merits of using deterministic approaches, such as the commonly used iterative reconvolution method, and probabilistic approaches, such as the exponential series method (ESM) and recently proposed basis pursuit denoising (BPDN, compressed sensing) method. A deconvolution approach, based on the process of simulated annealing (SA), was also reviewed. Generally speaking, all of the methods performed well and could accurately recover the simulated decay parameters. The fits obtained from the real decays were also fairly consistent between analysis methods. The accuracy of the probabilistic approaches was particularly noteworthy because of their lack of constraint to a particular functional form. The only considerable shortcoming that was observed was the significant error the SA deconvolution approach suffered when a short lifetime component was present in the system. This issue would need to be resolved before SA deconvolution could be considered to be a credible analysis method; however, as has been previously noted, the error observed may have been a manifestation of the use of inappropriate input parameters. The main outcome of this review of analytical techniques was that the various methods had different strengths and weaknesses and that using a combination of methods may improve the robustness of the conclusions drawn from the results. A possible strategy would be to use deterministic approaches to refine a model that was implicated by probabilistic approaches. The analysis code that was written for this investigation should be of significant value for future studies. Although some improvements are required to optimise their performance, the complementary nature of the analysis techniques could help to provide crucial new insights into systems where it has so far been challenging to determine a reliable physical model.

The conformational stability of 2AP-containing dinucleotides was investigated by performing computational calculations at the M06-2X/6-31+G(d) level of theory.

Previous time-resolved fluorescence measurements had shown that there is remarkable similarity between the photophysical behaviour of 2AP incorporated into dinucleotide and duplex constructs. Typically this observation has been concluded to imply that the dinucleotide system accesses duplex-like conformations with similar probability; however, this has always been based on limited direct evidence. The results presented here give much greater weight to this interpretation. In particular, dinucleotides have been shown to form stable, efficiently base-stacked states similar to those observed in the highly organised duplex structure. In both systems, this base-stacked conformation could potentially facilitate rapid charge transfer and, in turn, account for the short lifetime component observed in the fluorescence decay. Despite being structurally similar, dinucleotides were generally found to have a significantly larger twist angle than the value of 36° typically quoted for DNA. The same effect was seen in both 2AP dinucleotides and equivalent adenine dinucleotides and so could not simply be attributed to the use of an unnatural base. Comparison with the results obtained for the dimer structures (where there was no constraint from the backbone) seemed to suggest that the large twist values were due to the preferred stacking orientation of the bases. The study also found a stable, open dinucleotide which, like the base-flipped state in a duplex, would seem like a plausible source for the longest lifetime component of the fluorescence decay. 2AP appeared to be liberated from intramolecular interactions in the open dinucleotide conformation and so would be expected to behave as if it were essentially free in solution. Further investigation is required to determine the source of stability in the open structure; intuitively it is difficult to see why such a structure would be a minimum of energy. It is possible that the structure has optimised bond angles and dihedrals that would otherwise be constrained in a more condensed form but, if this were the case, it would be expected that there would be many open conformations of similar energy. Regardless of the cause of its stability, the ability to maintain an open (or base-flipped) conformation could be of importance for enzymatic interactions with DNA. For instance, does an enzyme need to flip and capture a target base or could it simply rely on the intrinsic stability of the open state? Comparison of analogous adenine and 2AP dinucleotides has shown that 2AP can mimic natural base-stacking well; however, there were slight differences that might help to explain experimental results. Specifically, formation of an H-bond between the 2AP amino group and the sugar of the neighbouring base may inhibit base-pairing with thymine in a duplex and, subsequently, be the cause of the increased imino proton exchange rate observed in NMR experiments. Further

investigation is required to validate this speculative hypothesis though. Calculations performed on nucleobase dimer structures highlighted the fact that the sugar-phosphate backbone may influence the stability of the base-stacking geometry. This is an important observation as the backbone is often omitted to reduce computational cost. Excluding PCM solvent during calculations appeared to have little effect on the conformation of the dinucleotides; however, these calculations were likely to have been significantly biased by using the optimised aqueous structures for the initial geometry. There is considerable scope for further investigation using the methodologies that have been developed throughout the course of this study. Collaboration with Dr. Tanja van Mourik and Leo Holroyd from the University of St Andrews is ongoing and has already produced further insights into the conformational landscape of nucleic acids. Of particular note has been the determination of stable dinucleotide structures that are even more similar to canonical, B-form DNA than those observed in the study presented here. Unfortunately, these new results were obtained too late to discuss in detail within this thesis but it would appear that hyper-rotation of the twist angle is not essential for dinucleotide stability and its manifestation was likely a consequence of the initial starting geometry used. There has also been considerable effort towards gaining insight into the base-flipping process of nucleic acids by investigating the stability of structures between fully stacked and open states; this may provide conformations that could explain the intermediate lifetime components observed in time-resolved fluorescence measurements. Beyond the studies already mentioned, it is foreseeable that there would be great benefit in investigating larger constructs, which more accurately mimic the conformational constraints found within the duplex. Extending the size of the system would incur significant computational cost but might be necessary to realise a true representation of DNA structure and dynamics.

CyDNA, a densely labelled and highly fluorescent DNA-based biopolymer, has been investigated to determine its suitability for use in fluorescence microscopy applications. Photophysical studies have identified dye-dye interactions, such as the formation of H-aggregates and resonance energy transfer, as a source of significant quenching in this system, particularly at high label density. This compromises performance for potential applications such as fluorescence *in situ* hybridisation (FISH) and in microarray and microfluidic devices; however, it may be possible to reduce the severity of these detrimental quenching effects by optimising the design of the template used during the PCR-mediated synthesis process. Reversible fluorescence

photoswitching was introduced into CyDNA by utilising the desirable properties of the Cy3-Cy5 pair. Two separate strategies produced substrates of slightly different character. Fluorescence photoswitching was shown to be reliable at ensemble and single-molecule level for both variants but, depending on the application, one or the other might be preferred. Discrimination and image contrast enhancement of photoswitching CyDNA from its non-switching variant was found to be possible through the use of optical lock-in detection imaging. This could enable high contrast imaging of well-defined targets at low probe concentrations. Importantly, these developments also offer the prospect of improved and sequence-specific super-resolution microscopy of DNA, which could provide crucial new insights into its 3D nanoscale structure in cellular environments.

Over the last few decades there have been great advances in the understanding of DNA. In particular, it has become clear that conformational dynamics play an important role in determining its functional behaviour. Given the fundamental importance of DNA to life, it is essential that this drive for knowledge continues and that the remaining secrets of this fascinating and remarkable biopolymer are uncovered.

Publications

Reversible Fluorescence Photoswitching in DNA

D. A. Smith, P. Holliger, and C. Flors, *J. Phys. Chem. B*, 2012, **116**, 10290–3.

DOI: 10.1021/jp3056834

URL: http://pubs.acs.org/doi/abs/10.1021/jp3056834

Enzyme-Promoted Base Flipping Controls DNA Methylation Fidelity

D. M. Matje, H. Zhou, D. A. Smith, R. K. Neely, D. T. F. Dryden, A. C. Jones, F. W. Dahlquist, and N. O. Reich, *Biochemistry*, 2013, **52**, 1677–85.

DOI: 10.1021/bi3012912

URL: http://pubs.acs.org/doi/abs/10.1021/bi3012912

Lectures, Courses, and Training

Laser Safety Course

Safety Risk Assessment Lecture

General Safety Lecture

LaTeX for Beginners

Managing your PhD (Chemistry)

Introduction to Computational Chemistry (Scottish Universities Physics Alliance)

Biophysical Chemistry Lecture Course

Postgraduate Essentials

School of Chemistry Colloquia

Safety Exam (Melbourne, Australia)

EndNote (Windows): Introductory Workshop (Melbourne, Australia)

Induction and Safety Program (Melbourne, Australia)

MATLAB Fundamentals (Melbourne, Australia)

IP and Commercialisation (Melbourne, Australia)

Publishing Strategically in the Sciences (Melbourne, Australia)

EaStChem PhD Thesis Workshop

Gaussian Workshop

Effective Writing: The Writing Process

Conferences and Meetings

Physical Chemistry Joint Meeting with Heriot-Watt University, Firbush Point, Killin, May 2011.

12th Conference on Methods and Applications of Fluorescence: Spectroscopy, Imaging and Probes, Strasbourg, France, September 2011.

BioPhysChem2011, Wollongong, Australia, December 2011.

26th International Conference on Photochemistry, Leuven, Belgium, July 2013.

13th Conference on Methods and Applications of Fluorescence: Spectroscopy, Imaging and Probes, Genoa, Italy, September 2013.

Edinburgh 300: Cradle of Chemistry, The Royal Society of Edinburgh, Edinburgh, October 2013

Physical Chemistry Section Meetings, Firbush Point, Killin.